



Facultad de Psicología

Tesis Doctoral:

**EL ANÁLISIS FACTORIAL DE VARIABLES IPSATIZADAS
COMO ESTRATEGIA PARA EVITAR EL IMPACTO DEL SESGO
DE RESPUESTA EN ESTUDIOS DE INVARIANZA FACTORIAL.**

Realizada por: Carlos Calderón Carvajal

Dirigida por: Carmen Ximénez Gómez

Departamento de Psicología Social y Metodología

Madrid, Abril de 2014.

**EL ANÁLISIS FACTORIAL DE VARIABLES IPSATIZADAS COMO
ESTRATEGIA PARA EVITAR EL IMPACTO DEL SESGO DE
RESPUESTA EN ESTUDIOS DE INVARIANZA FACTORIAL.**

Por

CARLOS CALDERÓN CARVAJAL

Tesis para optar al grado académico de Doctor en Psicología

Universidad Autónoma de Madrid

Facultad de Psicología

Madrid, España

Abril de 2014

Copyright © Carlos Calderón, 2014

AGRADECIMIENTOS

Quiero empezar los agradecimientos pidiendo perdón. Perdonadme todos quienes, por falta de espacio o memoria, no podré mencionar en este apartado. Como si esto exculpara a mi conciencia, continúo.

En primer lugar, quiero agradecer a la Comisión Nacional de Investigación Científica y Tecnológica, CONICYT, que me ha otorgado la beca, sin la cual no hubiese sido posible realizar mis estudios de doctorado.

A la Escuela de Psicología de la Universidad Católica del Norte, con quienes nunca hemos perdido el vínculo. A Paula, Jaime, Mariana, Jimena, Alfonso y Mónica, quienes siempre han creído en mí. En especial a Manuel, quien me inoculó el virus de la investigación y me inició en mis primeros pasos en la academia.

Al Departamento de Psicología Social y Metodología de la Universidad Autónoma de Madrid. Rodearse y aprender de profesores e investigadores de tanto nivel, no hace más que sentirme orgulloso de haber pertenecido y haber realizado mis estudios doctorales en esta casa de estudios.

A mi familia, quienes han tenido que soportar mi lejanía. Miriam, Alejandra, Juan, Jaime, Tita, Rosa, Isabel, siempre pendientes y al tanto. A María, mi madre, quien me crió y me hizo ser la persona que soy. A mi Padre, quien no pudo ver a quien he llegado a ser.

A los amigos que dejamos en Madrid. Los compañeros del programa, de quienes aprendí mucho más que metodología. A María, Jacinta y Gerardo, quienes nos recibieron con los brazos abiertos, y hoy nos sentimos parte de su familia. A Elena, Giulia, Blanca,

Mica, Luis, Juanki, Assia, Juanjo, Rodrigo, Andrés, Julio, Amadeu, Jesús, Jorge, Bere y cuantos otros, quienes nos abrieron sus hogares, y nos hicieron sentir como si fueran también los nuestros. Al “Botafumeiro”, la continuación del salón de nuestra casa.

A mi directora de esta tesis, Carmen Ximénez, quien a pesar de las dificultades, siempre tuvo la paciencia y dedicación para corregir mis barbaridades. Una persona de una bondad increíble para ofrecer sus conocimientos. De dedicación y esmero inagotables. Sin ella este trabajo no hubiese sido el mismo, quizá nunca acabado. Gracias por creer en mí y en mi trabajo, incluso cuando yo no lo hacía.

Finalmente a la persona más importante en mi vida. A la mujer que me acompaña, me soporta y me levante para continuar por este derrotero. Quien amo por sobre toda las cosas. Mi hogar esté donde esté. Mi refugio y mi calma. Con quien estuvimos juntos en esta loca aventura, y espero continuar muchísimas más. Quien me da razón de seguir luchando. Quien incluso ha tenido que sufrir el martirio de tener que leer, corregir y comentar esta tesis. Quien me ha enseñado a vivir la vida: perdóname por no ser tan bueno como te mereces.

Madrid, Otoño de 2014

A Mono, a quien le debo la vida.

... y a Gerardo, quien se fue debiéndome un par de cervezas.

EL ANÁLISIS FACTORIAL DE VARIABLES IPSATIZADAS COMO ESTRATEGIA PARA EVITAR EL IMPACTO DEL SESGO DE RESPUESTA EN ESTUDIOS DE INVARIANZA FACTORIAL

RESUMEN

La Equivalencia/Invarianza de Medida (E/IM) de un test, es una condición necesaria para la adecuada comparación entre diferentes grupos. La principal herramienta para la evaluación de esta propiedad es el Análisis Factorial Confirmatorio Multi-grupo (AFCM). Pese a las ventajas de esta herramienta, a menudo los investigadores se enfrentan al problema del *sesgo de respuesta*, el cual puede afectar a los resultados de esta clase de análisis. Una de las propuestas para tratamiento del *sesgo de respuesta* es la transformación *ipsativa* o *ipsatización*. Los escasos estudios existentes que han analizado el funcionamiento del AFC de variables *ipsatizadas*, han mostrado una pérdida significativa de la sensibilidad de los índices de bondad de ajuste para detectar el error de especificación. A pesar de la potencial ventaja de la *ipsatización*, hasta el momento no existen estudios que evalúen la efectividad de esta estrategia para eliminar el *sesgo de respuesta* y las condiciones óptimas para llevar a cabo un AFCM con variables *ipsatizadas*.

El objetivo de esta tesis es evaluar el funcionamiento de la *ipsatización* de variables en el tratamiento del *sesgo de respuesta* y las condiciones óptimas para llevar a cabo un AFCM de variables *ipsatizadas*, especialmente en ausencia de invarianza factorial. Los resultados son sugerentes con respecto a la efectividad de la *ipsatización* para evitar efecto del *sesgo de respuesta* y ofrecen algunas recomendaciones prácticas para llevar a cabo un AFCM de variables *ipsatizadas* con ciertas garantías.

FACTOR ANALYSIS OF IPSATIZED VARIABLES AS AN APPROACH TO REMOVE RESPONSE BIAS IN MEASUREMENT INVARIANCE STUDIES

ABSTRACT

The Measurement Equivalence/Invariance of a test is a necessary condition to the accurate comparison across different groups. The main approach for the assessment of this property is the Multiple-group Confirmatory Factor Analysis (MCFA). Despite the advantage of this procedure, often researchers face the problem of the *response bias*, that may affect the results of such analysis. One proposal for the treatment of *response bias* is the *ipsative transformation* or *ipsatization*. The few studies that have analyzed the performance of CFA *ipsatized* variables have shown a significant lack of sensitivity of the goodness of fit indexes to detect the misspecification. Despite the potential advantage of *ipsatization*, so far there are no studies that evaluate the effectiveness of this strategy to eliminate *response bias* and the optimal conditions for carrying out a MCFA with *ipsatized* variables.

The aim of this thesis is to evaluate the performance of the *ipsatization* of variables in the treatment of *response bias* and the optimal conditions for conducting a MCFA of *ipsatized* variables, when there is lack of factorial invariance. The results regarding the effectiveness of the *ipsatization* to avoid the effect of *response bias* are challenging and offer some practical advice for conducting a MCFA of *ipsatized* variables with certain guarantees.

ÍNDICE

INTRODUCCIÓN.....	1
CAPITULO I: ASPECTOS TEÓRICOS.....	11
1.1. La equivalencia/invarianza de medida (E/IM).....	11
1.2. El Análisis Factorial Confirmatorio Multi-grupo (AFCM).....	15
1.3. El problema del sesgo de respuesta.....	24
1.4. El análisis factorial de datos <i>ipsatizados</i>	27
1.4.1. El problema del análisis de datos <i>ipsatizados</i>	27
1.4.2. El AFC de daros <i>ipsatizados</i>	29
1.4.3. Implementación del método DE en un modelo SEM.....	34
CAPITULO II: LOS TEST IPSATIVOS DE RESPUESTA FORZADA.....	39
2.1. El formato de respuesta forzada.....	41
2.2. El modelo Thurstoniano.....	42
2.3. Análisis factorial de ítems de respuesta forzada.....	43
CAPITULO III: ANTECEDENTES Y ESTADO ACTUAL DEL PROBLEMA.....	53
3.1. Funcionamiento del AFC de variables ipsatizadas.....	53
3.2. El error de especificación en el AFC de variables ipsatizadas.....	55
3.3. Sensibilidad de los índices de ajuste a la falta invarianza factorial.....	60
CAPITULO IV: OBJETIVOS DE LA TESIS.....	65
4.1. Objetivos.....	65
4.2. Plan de trabajo.....	68
4.3. Hipótesis.....	71

CAPITULO V: ESTUDIOS DE SIMULACIÓN.....	75
5.1. Estudio 1: Sensibilidad de los índices de ajuste a la invarianza configural en el AFCM de datos <i>ipsatizados</i>	75
5.2. Estudio 2: Sensibilidad de los índices de ajuste a la invarianza métrica en el AFCM de datos <i>ipsatizados</i>	101
5.3. Síntesis general y recomendaciones prácticas de los estudios de simulación.....	123
CAPÍTULO VI: ESTUDIO EMPÍRICO.....	129
6.1. Método.....	129
6.2. Resultados.....	133
6.3. Conclusiones.....	138
CAPÍTULO VII: CONCLUSIONES Y COMENTARIOS FINALES.....	141
REFERENCIAS.....	153
ANEXOS.....	163
Terminología.....	165
Anexo 1: Sintaxis EQS para la estimación del modelo de 3 factores de 3 ítems por factor....	167
Anexo 2: Sintaxis EQS para la estimación del modelo del Inventario de Estilos de Aprendizaje de KOLB (1985). Versión reducida ipsatizada.....	169
Anexo 3: Sintaxis MPLUS para la estimación del modelo TRI multidimensional de Brown y Maydeu-Olivares (2010) del Inventario de Estilos de Aprendizaje de KOLB (1985) para la versión ipsatizada.....	173
Anexo 4: Parámetros teóricos utilizados en el estudio 1	177
Anexo 5: Parámetros teóricos utilizados en las condiciones de invarianza del estudio 2	179
Anexo 6: Instrumento utilizado en el estudio empírico. Inventario de Estilos de Aprendizaje de Kolb (1985). Versión reducida.....	181

INTRODUCCIÓN

Un problema frecuente en la práctica investigadora es el *sesgo de respuesta* o la tendencia sistemática a responder un test independientemente del contenido y de la puntuación verdadera del individuo. El sesgo de respuesta afecta tanto a la fiabilidad como a la validez de las medidas, y junto con ello, a las inferencias hechas a partir de las puntuaciones en el test. El problema cobra especial importancia cuando estas inferencias involucran comparaciones entre grupos. Una forma de tratar el *sesgo de respuesta* es mediante la *ipsatización* de las puntuaciones. En modelos correctamente especificados, los métodos de Análisis Factorial Confirmatorio (AFC) de datos *ipsatizados* han demostrado recuperar la estructura factorial de manera satisfactoria, mientras que presentan dificultades para detectar la falta de ajuste cuando existen errores de especificación (Ximénez y Calderón, 2012).

A pesar de la potencial ventaja que ofrece la *ipsatización* de variables para evitar el impacto del sesgo de respuesta, hasta la fecha no hay estudios que evalúen el funcionamiento del AFC de datos *ipsatizados* en el análisis de la invarianza factorial. Es por ello que el objetivo del presente trabajo es examinar el funcionamiento del AFC de datos *ipsatizados* en la evaluación de la invarianza factorial entre grupos. Ya que el principal problema que presentan esta clase de modelos es la sensibilidad de las medidas de bondad de ajuste, nuestro trabajo se centra en evaluar la efectividad de algunos de los índices de bondad de ajuste más utilizados para detectar la falta de invarianza factorial en el contexto del análisis de la invarianza de medida.

En ciencias sociales es frecuente que los investigadores estén interesados en detectar diferencias entre distintos grupos. Para evaluar adecuadamente estas diferencias es esencial que las medidas utilizadas tengan las mismas propiedades psicométricas en los grupos que se comparen. En otras palabras, para que estas comparaciones sean válidas, es necesario que las medidas sean equivalentes o invariantes entre los grupos (Horn y McArdle, 1992).

La Equivalencia/Invarianza de Medida (E/IM, en adelante) es la propiedad que permite a un instrumento obtener medidas del mismo atributo en distintos grupos o situaciones (Horn y McArdle, 1992). El incumplimiento de la E/IM puede conducir a sesgos importantes en los resultados e inferencias hechas por los investigadores (Borsboom, 2006; Meredith y Teresi, 2006). Esto es debido a que la falta de E/IM hace que las medidas no tengan el mismo significado en los grupos que se comparan.

Por ejemplo, en un estudio pionero, Millsap y Kwok (2004) estudiaron el efecto de la falta de equivalencia en medidas utilizadas para la selección de sujetos pertenecientes a diferentes grupos. Los autores demostraron que pequeñas diferencias en la estructura factorial del instrumento pueden conducir a errores en la selección de los sujetos pertenecientes a distintos grupos.

Una técnica comúnmente utilizada para evaluar la E/IM es el Análisis Factorial Confirmatorio Multi-grupo (AFCM, en adelante). Esta técnica permite evaluar la estabilidad de la estructura interna de un conjunto de datos en distintas condiciones. Estas condiciones consideran, entre otras: la estabilidad entre distintos grupos o poblaciones (Labouvie y Ruetsch, 1995; Little, 1997), a través del tiempo

(Golembiewski, Billingsley y Yeager, 1975), en distintos evaluadores (Facteau y Craig, 2001), o formatos de administración (Taris, Bok y Meijer, 1998).

Aunque el procedimiento particular para llevar a cabo un AFCM dependerá de los objetivos del investigador, la estrategia más extendida contempla un proceso compuesto de varias etapas, en las cuales se imponen restricciones de manera secuencial y jerárquica. El procedimiento suele iniciarse con el contraste de igualdad de los patrones de los parámetros o forma de la estructura factorial (modelo base o nivel configural), hasta llegar al nivel más restrictivo que suele considerar la igualdad en los errores de medida (invarianza estricta). A pesar de la existencia de distintos niveles de invarianza, existe un consenso general en que la igualdad de las saturaciones factoriales (nivel de invarianza métrico) es condición suficiente para determinar la E/IM, cuando el foco del análisis está puesto en la validez de constructo (Dimitrov, 2010).

Pese a las ventajas que posee el AFCM para determinar la E/IM, los investigadores a menudo se enfrentan al problema del *sesgo de respuesta*. Este problema tiene una larga tradición en la investigación social, ya que ha mostrado afectar a la fiabilidad y la validez de las medidas (Cheung y Rensvold, 1999; Marquis, Marquis y Polich, 1986), y con ello, a las inferencias establecidas a partir de los resultados del AFCM.

Desde un punto de vista metodológico, los intentos por evitar este problema se han centrado principalmente en dos tipos de enfoques. El primero de ellos, el enfoque instrumental, se refiere al diseño del experimento y/o de los instrumentos de medida. Aquí la estrategia se centra en la forma de obtener los datos. Un ejemplo de este tipo

de enfoque es el uso de los test de *respuesta forzada*. En esta clase de test, a los entrevistados se les presentan preguntas con alternativas con un grado de deseabilidad similar. Ya que el entrevistado está obligado a priorizar unas alternativas sobre otras, este formato de respuesta contribuye a evitar el efecto de la *deseabilidad social* o la *acquiescencia*, que son las fuentes más frecuentes del *sesgo de respuesta* (Cheung y Chan, 2002; Cheung y Rensvold, 2002).

El segundo, el enfoque estadístico, foco principal de esta tesis, busca eliminar el *sesgo de respuesta* por medio del tratamiento de los datos una vez que estos han sido obtenidos. Una de estas estrategias consiste en la estandarización de las puntuaciones. Aunque existen varios tipos de estandarización, el propósito general de estas estrategias es corregir la tendencia en los datos debido a la presencia de *sesgo de respuesta* (Fisher, 2004).

Cheung y Chan (2002) proponen un tipo de estandarización llamado *transformación ipsativa* o *ipsatización* (Hicks, 1970). Esta transformación corresponde a una estandarización intra-sujeto, ya que las puntuaciones se ajustan con respecto a la media de cada individuo (Cunningham, Cunningham y Green, 1977; Hofstede, 1980).

A pesar de la ventaja potencial para eliminar el sesgo de respuesta, este tipo de estandarización genera datos con propiedades *ipsativas*. Estas propiedades hacen desaconsejable el uso de técnicas convencionales de análisis de datos, ya que los datos *ipsativos* incumplen una serie de asunciones necesarias desde el punto de vista analítico (Aitchison, 1986). Concretamente, en el estudio de la estructura de covarianzas, los datos *ipsativos* generan matrices de covarianzas singulares, no

invertibles, las cuales no pueden ser factorizadas por medio de los métodos convencionales (Chan y Bentler, 1993).

Para solucionar este problema, Chan y Bentler (1993), y posteriormente, Cheung (2004), desarrollaron un método de factorización para variables *ipsativas*. En estos trabajos, los autores demuestran que el modelo de AFC para variables *ipsativas* logra recuperar el modelo *pre-ipsativo* o *normativo* de manera satisfactoria. A pesar de ello, apenas existen estudios que evalúen las condiciones óptimas de aplicación de estos procedimientos, y los pocos que hay, evalúan un número reducido de condiciones y sólo consideran condiciones en las cuales los modelos están correctamente especificados (Chan y Bentler, 1996; Chan, 2003; Cheung, 2004).

Recientemente, Ximénez y Calderón (2012) han demostrado que, a pesar de que el AFC de variables *ipsatizadas* funciona adecuadamente al recuperar modelos correctamente especificados, éste empeora considerablemente bajo errores de especificación. En concreto, los índices de ajuste pierden sensibilidad, teniendo serias dificultades para detectar modelos incorrectamente especificados.

Pese a la importancia del AFCM para evaluar la E/IM y de la potencial ventaja de la *ipsatización* para eliminar el *sesgo de respuesta*, no existen trabajos que evalúen el funcionamiento del AFC de variables *ipsatizadas* en el análisis de la invarianza factorial. La literatura previa, los pocos trabajos publicados (véase Chan, 2003; Chan y Bentler, 1996; Cheung, 2004; Ximénez y Calderón; 2012) sólo ha examinado el funcionamiento del AFC de datos *ipsatizados* en un número muy reducido de condiciones, y ninguno de estos trabajos ha evaluado el AFCM con datos *ipsatizados*. Es por ello que creemos que es de suma importancia evaluar tanto la

eficacia de la *ipsatización* para eliminar el *sesgo de respuesta*, como el funcionamiento general del AFCM con datos *ipsatizados*, y en particular, determinar la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza en el contexto de E/IM con datos *ipsatizados*.

El estudio del funcionamiento del AFCM con variables *ipsatizadas*, y el conocimiento de las condiciones óptimas para llevar a cabo esta clase de análisis, puede contribuir a la utilización de este tipo de técnica de estandarización como estrategia de tratamiento del *sesgo de respuesta* en el análisis de la E/IM.

Es por ello que el objetivo de esta tesis es evaluar el funcionamiento del AFC para variables *ipsatizadas* en el análisis de invarianza factorial entre grupos. Concretamente, nuestro interés se centra en la evaluación de los índices de ajuste para detectar la falta de invarianza entre grupos, ya que la literatura previa ha demostrado que la sensibilidad de las medidas de bondad de ajuste se ve especialmente afectada en el AFC de variables *ipsatizadas*.

Sólo nos centraremos en el estudio de la invarianza a nivel configural y métrico, ya que el cumplimiento de la invarianza en estos dos niveles es condición suficiente para asegurar la E/IM, en el contexto de la validez de constructo. Así mismo, también consideraremos la evaluación de esta técnica con datos reales a través de la comparación de su funcionamiento con otra de las estrategias para abordar el *sesgo de respuestas*: los test de *respuesta forzada*.

El trabajo está organizado en siete capítulos. En el capítulo I se exponen los aspectos teóricos fundamentales para la comprensión y planteamiento de nuestro problema de investigación. Este capítulo está organizado en cuatro secciones. En la primera de ellas se plantea la importancia de la E/IM en estudios que incorporan el análisis de dos o más grupos, así como la importancia del AFCM como principal estrategia para la evaluación de la E/IM. En la segunda sección se presenta una breve descripción del fundamento analítico del AFC y su extensión al análisis de la invarianza factorial de dos o más grupos, así como el procedimiento general de implementación y la descripción de cada uno de los niveles de invarianza de medida. En la tercera sección se presenta una definición del sesgo de respuesta y sus diferentes tipos. Así mismo, se presenta su definición desde el punto de vista formal, y la fundamentación analítica de cómo la *transformación ipsativa* o *ipsatización*, logra eliminar el *sesgo de respuesta* de un conjunto de datos. Finalmente, la cuarta sección presenta los problemas de los datos con propiedades *ipsativas* en general, y específicamente, el problema analítico en el contexto del análisis de la estructura de covarianzas con datos *ipsatizados*. Junto con ello se describe formalmente el modelo de AFC de variable *ipsatizadas* y su implementación en un programa SEM.

El capítulo II presenta una de las alternativas al tratamiento instrumental del *sesgo de respuesta*: los test de *respuesta forzada*, uno de los tipos de test más utilizados en el contexto de esta clase de enfoque. Se describen las características de los test de *respuesta forzada*, así como la naturaleza analítica de los datos que producen. Finalmente, se presenta uno de los modelos de AFC de datos obtenidos a partir de los test de *respuesta forzada*, así como algunos de los pocos factores que se

han estudiado y que afectan a los resultados de esta clase de análisis. Utilizaremos este formato de test en uno de nuestros estudios para comparar su funcionamiento con el de la *ipsatización* en el tratamiento del *sesgo de respuesta*.

En el capítulo III se revisan los principales hallazgos relacionados con el AFC de datos *ipsatizados* y el AFCM. Debido a que no existen estudios previos que evalúen el funcionamiento del AFCM con variables *ipsatizadas*, la revisión se centra en los hallazgos relacionados con el AFCM de variables *normativas*. Así mismo, se revisan los estudios previos que abordan el estudio del AFC con variables *ipsatizadas*, y concretamente, los problemas de sensibilidad de los índices de ajuste en la detección del error de especificación.

El capítulo IV describe en detalle los objetivos de la tesis, así como las hipótesis planteadas en relación a la evidencia presente en la literatura.

En el capítulo V se presentan dos estudios de simulación que pretenden evaluar la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza. El estudio 1 está centrado en la detección de la invarianza a nivel configural. En base a estudios anteriores, hemos introducido la falta de invarianza a través de la omisión de un factor en uno de los grupos (*infra-factorización*). El estudio 2 estudia la falta de invarianza a nivel métrico. En este estudio hemos introducido la falta de invarianza a través de las diferencias en las saturaciones factoriales en grupos con la misma estructura factorial. En ambos estudios se presentan recomendaciones prácticas para los investigadores interesados en realizar análisis de invarianza para datos *ipsatizados*.

El capítulo VI presenta los resultados de un estudio empírico que pretende evaluar el funcionamiento del AFCM con datos *ipsatizados* y su implementación en un programa SEM. Para evaluar el funcionamiento de la *ipsatización* como estrategia para eliminar el sesgo de respuesta, comparamos los resultados con los datos obtenidos con un test de *respuesta forzada*, la versión reducida del “Inventario de Estilos de Aprendizaje”, LSI de Kolb (1985), uno de los test más utilizados en el ámbito educativo. Este test ha sido presentado a los encuestados en dos formatos, el formato original (formato de *respuesta forzada*), y la versión *normativa* de mismo test (escala Likert). Posteriormente hemos *ipsatizado* los datos *normativos* obteniendo tres conjuntos de datos sobre la misma muestra: *normativos*, *ipsatizados* e *ipsativos* (o de *respuesta forzada*). Finalmente, comparamos los resultados obtenidos al ajustar los tres conjunto de datos.

Por último, en el capítulo VII, se presentan las conclusiones generales y comentarios finales acerca de los resultados obtenidos en los tres estudios, así como recomendaciones prácticas para los investigadores interesados en trabajar con datos *ipsatizados*. También se comentan algunas propuestas acerca de las líneas de investigación abiertas en el estudio del AFCM de variables *ipsatizadas*.

Finalmente, se incluyen las referencias así como los anexos donde se describen algunos detalles del diseño de los estudios de simulación llevados a cabo, las sintaxis implementadas en el programa informático que hemos utilizado para el análisis de variables *ipsatizadas*, el cuestionario utilizado en el estudio empírico, así como un glosario de terminología con los principales símbolos y abreviaturas presentes en el texto.

Para el lector familiarizado con el AFCM y su implementación, recomendamos omitir la primera y segunda sección del capítulo I que trata el fundamento analítico y el procedimiento de aplicación de esta técnica. Para lectores avanzados o para quienes desean interiorizarse en el fundamento analítico y aplicación del AFC con datos *ipsatizados*, se recomienda la revisión en detalle de la sección 4 del capítulo II. Así mismo, para aquellos familiarizados con los test con formato de *respuesta forzada* y que estén interesados en la propuesta de AFC de este tipo de datos, así como su implementación en un programa informático, recomendamos la revisión detallada del capítulo II.

CAPITULO I

ASPECTOS TEÓRICOS

Este capítulo considera la revisión de aspectos teóricos fundamentales de la utilización de la *ipsatización* en el estudio de la E/IM, en el marco del modelo contemporáneo de validez (Messick, 1995). Ya que nuestro trabajo se centra en la utilización del AFCM como técnica para evaluar la E/IM, el objetivo del capítulo es realizar una descripción general de esta técnica aplicada a la evaluación de aspectos relacionados con la generalización de las propiedades psicométricas de una medida. Se presenta la definición conceptual y analítica del *sesgo de respuesta* y cómo este puede afectar a los resultados del análisis de la E/IM. Junto con ello, se revisa el fundamento analítico de cómo la *ipsatización* puede evitar su impacto sobre un conjunto de datos. Finalmente, se presentan los problemas analíticos del análisis de datos *ipsativos* en general y el procedimiento para analizar los datos *ipsatizados* en particular. En síntesis, este capítulo abarca los aspectos teóricos fundamentales para comprender la utilización de la *ipsatización* como estrategia para evitar el impacto del *sesgo de respuesta* en el análisis de la E/IM, y cómo deben analizarse un conjunto de datos *ipsatizados* a través del AFC.

1.1. LA EQUIVALENCIA O INVARIANZA DE MEDIDA

Desde una perspectiva contemporánea, se considera la validez como un juicio evaluativo integral, acerca del grado en el cual la evidencia empírica y los fundamentos teóricos, apoyan la conveniencia y pertinencia de las inferencias y

acciones basadas en los resultados obtenidos de la medida de un determinado *constructo* (Messick, 1989).

Esta perspectiva aborda la validez a partir de un “modelo unificado basado en el *constructo*” (*unified construct-based model*). Según este modelo, los distintos tipos de inferencias que se pueden hacer desde las puntuaciones de un test, requieren de diferentes tipos evidencias que apoyen su distintos usos (Messick, 1995).

Messick (1995) especifica seis aspectos de esta concepción unificada de validez de constructo, los cuales poseen un carácter tanto científico como ético. Estos aspectos consideran distintas fuentes de evidencia de validez: (1) *de contenido*, (2) *de aspectos sustantivos*, (3) *de estructura interna*, (4) *de generalización*, (5) *de evidencias externas* y (6) *de consecuencias éticas*. Dos de estos aspectos involucran la utilización del AFC en el contexto de la evaluación de la invarianza de medida; la evaluación de la *estructura interna* y la *capacidad de generalización* de sus propiedades psicométricas.

El *aspecto estructural* evalúa la coherencia entre el sistema de puntuación y la estructura interna del constructo de interés. La búsqueda de evidencias de aspectos estructurales se centra en determinar la consistencia entre el constructo subyacente y los indicadores por medio de los cuales es medido. La principal estrategia para recolectar evidencia de *aspectos estructurales* de validez es el AFC.

El contraste de aspectos estructurales a través del AFC en un grupo o muestra, no asegura que sus resultados puedan ser generalizables a otros grupos de sujetos. Para ello, es necesario determinar que el modelo factorial sea invariante a través de los distintos grupos o condiciones.

Los aspectos relacionados con la *generalización* examinan el grado en que las propiedades de las puntuaciones y sus interpretaciones pueden ser generalizables a distintos grupos, contextos y/o momentos en el tiempo. Cuando los procedimientos de validación involucran la comparación de distintos grupos, es importante asegurar que el instrumento de medición opere de la misma forma en los distintos grupos que se comparan. En términos generales, esta propiedad de un instrumento de medida se denomina Equivalencia y/o Invarianza de Medida (E/IM).

La E/IM puede ser considerada como el grado en que un instrumento puede obtener medidas equivalentes bajo diferentes condiciones (Drasgow, 1987; Horn y McArdle, 1992). Estas condiciones pueden incluir la estabilidad de medida a través del tiempo (Chan, 1998; Chan y Schmitt, 2000), a través de diferentes medios de administración (Chan y Schmitt, 1997; Ployhart, Weekley, Holtz y Kemp, 2003); o a través de distintas poblaciones o grupos, como también distintas culturas (Riodan y Vandenberg, 1994), género (Marsh, 1985), grupos etarios (Faction y Craig, 2001), entre otras.

La E/IM puede ser definida como una independencia condicional, ya que existe si la distribución condicional de la variable observada, dado el factor latente, es independiente de la pertenencia grupal (Mellenbergh, 1989). Formalmente la E/IM se puede mantener si y sólo si (Yoon y Millsap, 2007):

$$P(\mathbf{X} | \mathbf{W}, G) = P(\mathbf{X} | G), \quad [1.1]$$

donde \mathbf{X} es un vector de variables observadas, \mathbf{W} es un vector de variables latentes subyacente a \mathbf{X} , y G representa el grupo de pertenencia. Como condición necesaria y suficiente para la E/IM, la ecuación [1.1] establece que la probabilidad condicional de

\mathbf{X} dado \mathbf{W} es independiente de G . La E/IM existe cuando la relación de las variables observadas con sus correspondientes variables latentes es independiente de la pertenencia grupal (Mellenbergh, 1989; Meredith y Millsap, 1992; Millsap, 1995).

De acuerdo a esta definición, cuando la E/IM existe, las personas de distintos grupos, y que poseen el mismo nivel en el rasgo latente, tienen la misma probabilidad de obtener las mismas puntuaciones observadas, independientemente del grupo al que pertenecen.

En ausencia de E/IM, se corre el riesgo de cometer errores graves en las decisiones que se tomen a partir de las puntuaciones del instrumento, debido a que las diferencias entre los individuos pueden ser reflejo de la relación diferencial entre el constructo y la puntuación observada, la cual no sería la misma en todos los grupos (Raju, Laffitte y Byrne, 2002). Por tanto, en ausencia de E/IM es difícil de interpretar sustantivamente las diferencias en las puntuaciones medias observadas entre los grupos.

Para evitar consecuencias éticas y sociales indeseables, el proceso de medición debe evitar el efecto de variables irrelevantes, como puede ser la pertenencia a un grupo (Messick, 1989). A su vez, debe incorporar métodos que ayuden a determinar el grado en el cual las puntuaciones se ven influidas por tales variables (Haladyna y Downing, 2004).

1.2. ANÁLISIS FACTORIAL CONFIRMATORIO MULTI-GRUPO (AFCM)

Una forma de determinar la E/IM entre los grupos, es evaluando la estabilidad de la estructura teórica subyacente a través del estudio de la invarianza factorial usando el Análisis Factorial Confirmatorio Multi-grupo (AFCM). El AFCM permite contrastar la estabilidad de una estructura teórica hipotética a través de varios grupos (Alwin y Jackson, 1981). Esta aproximación permite la comparación de características especiales del modelo factorial. En la medida en que estas características sean equivalentes en los grupos, el investigador puede inferir la existencia de invarianza factorial, y junto con ello, la existencia de E/IM (French y Finch, 2006).

En el AFCM, el correspondiente modelo de medida queda representado como:

$$\mathbf{y}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\xi}_g + \boldsymbol{\varepsilon}_g, \quad [1.2]$$

donde \mathbf{y} es un vector de puntuaciones observadas de orden $p \times 1$, $\boldsymbol{\tau}$ es un vector de interceptos de orden $p \times 1$, $\boldsymbol{\Lambda}$ es una matriz de saturaciones factoriales de orden $p \times q$, $\boldsymbol{\xi}$ es el vector de factores comunes de orden $q \times 1$, $\boldsymbol{\varepsilon}$ es un vector de orden $p \times 1$ de errores de medida, y g indica el grupo específico del parámetro. Asumiendo $E(\boldsymbol{\xi}\boldsymbol{\varepsilon}')=0$, la matriz de covarianzas queda expresada como:

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Lambda}_g \boldsymbol{\Phi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Psi}_g, \quad [1.3]$$

donde $\boldsymbol{\Sigma}_g$ es la matriz de covarianzas de las p variables observadas del grupo g , $\boldsymbol{\Phi}_g$ la matriz de covarianzas de los q factores, y $\boldsymbol{\Psi}_g$ la matriz diagonal de los p de errores de medida.

Asumiendo que la media o valor esperado de los errores de medida es igual a cero en la ecuación [1.2], el valor esperado para y en cada grupo es:

$$E(y_g) = \tau_g + \Lambda_g \kappa_g, \quad [1.4]$$

donde κ_g es la media del factor en el grupo g . Es decir $E(\xi_g) = \kappa_g$.

El AFCM involucra el ajuste de una serie de modelos anidados de manera secuencial en dos o más muestras. Estos modelos consideran subconjuntos de parámetros, los cuales son restringidos y evaluados de manera secuencial y jerárquica. Aunque no necesariamente, estos subconjuntos de parámetros suelen estar conformados por las distintas matrices y vectores del modelo.

Existen varias formas de clasificar los distintos niveles de invarianza (véase por ejemplo, Horn y McArdle, 1992; Meredith, 1993). Desde el punto de vista de la constitución del modelo, es posible diferenciar tres grandes niveles; *invarianza configural*, *invarianza de medida* e *invarianza estructural* (Byrne, 2004; Byrne et al., 1989; Cheung y Rensvold, 2002; Drasgow y Kanfer, 1985; Jöreskog y Sörbom, 1979; Meredith, 1993; Meredith y Teresi, 2006).

Invarianza Configural.

El nivel más básico de la invarianza factorial es el nivel configural (también denominado *invarianza de forma*), que establece que el número de factores, así como los patrones de saturaciones factoriales en cada factor, sean los mismos en los distintos grupos. En este nivel no existen restricciones de igualdad en ninguno de los valores de los parámetros. También es conocido como modelo base (Byrne, 2004;

Byrne et al., 1989), a partir del cual se ajustan una serie de modelos jerárquicos o anidados, más restrictivos, con el objetivo de detectar la fuente de invarianza.

Invarianza de Medida

La invarianza de medida contempla tres niveles de invarianza: la *invarianza métrica* (igualdad de las saturaciones factoriales entre los grupos), la *invarianza escalar* (igualdad de los interceptos de los ítems entre los grupos), y la *invarianza de la unicidades* (igualdad de la varianzas/covarianzas de los ítems entre los grupos).

El primero de estos niveles, la *invarianza métrica*, contrasta que los valores de las saturaciones factoriales de cada variable observada en cada factor (λ_{ij}) sean las mismas a través de los grupos. También es denominado, “*invarianza de medida débil*” (Dimitrov, 2010), compara que las unidades de cambio entre las variables observadas y los factores sean los mismos entre los grupos. Cuando se asume la *invarianza métrica*, tanto las media de los factores latentes, como los interceptos de los ítems pueden ser distintos en los diferentes grupos. Siguiendo al ecuación [1.2], la invarianza métrica queda expresada como:

$$\mathbf{y}_g = \boldsymbol{\tau}_g + \boldsymbol{\Lambda}\boldsymbol{\xi}_g + \boldsymbol{\varepsilon}_g. \quad [1.5]$$

Nótese que la diferencia entre la ecuación [1.2] y [1.5], es que los parámetros de la matriz $\boldsymbol{\Lambda}$ son los mismos para los g grupos.

El siguiente nivel, la *invarianza de medida escalar*, requiere que los interceptos sean iguales entre los distintos grupos. Para ello debe ser previamente establecida la igualdad de las saturaciones factoriales. El cumplimiento de este nivel

indica que la media de los ítems es la misma entre los distintos grupos. Este se denomina también “*nivel de invarianza de medida fuerte*” (Dimitrov, 2010). Siguiendo las ecuaciones [1.4] y [1.5], en la *invarianza escalar* el modelo de medida queda expresado como:

$$y_g = \tau + \Lambda \xi_g + \varepsilon_g, \quad [1.6]$$

y el valor esperado de y es:

$$E(y_g) = \tau + \Lambda \kappa_g. \quad [1.7]$$

Aquí el vector τ de interceptos queda restringido a ser igual en los g grupos.

Por último, la *invarianza de medida de las unicidades*, requiere del cumplimiento de la *invarianza métrica* y la *invarianza escalar*. Contrasta que tanto las varianzas como las covarianzas de los errores de medida de cada ítem sean iguales entre los grupos. El cumplimiento de este nivel, también denominado “*invarianza de medida estricta*”, es evidencia de que los ítems han sido medidos con la misma precisión o fiabilidad. En el modelo de medida, el nivel de *invarianza de las unicidades* queda expresado como:

$$y_g = \tau + \Lambda \xi_g + \varepsilon. \quad [1.8]$$

En este caso el vector ε de errores de medida se restringe a ser igual en los g grupos.

Invarianza Estructural

La invarianza estructural hace referencia a la parte estructural del modelo, es decir, a la igualdad de las varianzas y covarianzas de los factores (ϕ_{ii} y ϕ_{ij}). El

establecimiento de la invarianza estructural dependerá de si la variabilidad de los factores latentes y/o las relaciones entre los factores se consideran aspectos relevantes para la generalización de la validez de la medida. Es importante considerar que las comparaciones de las varianzas y/o covarianzas entre los factores sólo se pueden llevar a cabo si previamente se ha asegurado la invarianza de las saturaciones factoriales.

El procedimiento general para evaluar la invarianza factorial está descrito en una serie de fuentes (véase por ejemplo, Byrne, 2004; Byrne y Steward, 2006; Vandenberg y Lance, 2000). Globalmente, este procedimiento considera varias etapas. Un paso previo puede consistir en la aplicación de un *test ómnibus* de igualdad de matrices de covarianzas. Aunque para algunos autores puede ser suficiente para concluir sobre la E/IM (Jöreskog, 1971; Jöreskog y Sörbom, 1989), este test ha sido criticado por otros autores, puesto que ha mostrado inconsistencia en relación a otros test de invarianza factorial, por ejemplo, con el test de igualdad de conjuntos de parámetros en particular (Byrne, 1998; Raju, Laffite y Byrne, 2002),

Al margen de la prueba *ómnibus*, el primer paso consiste en el ajuste del nivel de invarianza configural. Como hemos mencionado anteriormente, en este nivel se examinan los grupos simultáneamente, manteniendo invariante sólo los patrones factoriales. Este modelo base, de idénticos patrones factoriales, provee un estadístico χ^2 que refleja el ajuste del modelo para los parámetros estimados separadamente.

A continuación se lleva a cabo el test de invarianza de las saturaciones factoriales, comparando el modelo base con un modelo idéntico, con la diferencia de

que la matriz de saturaciones factoriales se restringe a ser igual en todos los grupos (es decir, $\Lambda_g = \Lambda$). La diferencia entre el modelo base y el modelo más restrictivo se evalúa a través de la diferencia en el estadístico χ^2 , con grados de libertad igual al número de parámetros restringidos.

Si el propósito del AFCM es la evaluación de la E/IM, las siguientes dos etapas considerarán las restricciones de los parámetros pertenecientes a la invarianza de medida, a saber; la restricción en el vector de interceptos ($\tau_g = \tau$) y la evaluación de la invarianza de medida que contrasta la igualdad de las matrices de errores de medida ($\Psi_g = \Psi$).

Esta aproximación de etapas sucesivas en el AFCM se denomina procedimiento “*hacia adelante*” (*forward*). Bajo este enfoque, las evaluaciones de los distintos niveles de invarianza se organizan en una secuencia lógica de los modelos anidados, los cuales se establecen de una manera cada vez más restrictiva, para detectar la fuente de invarianza. En cada etapa se imponen nuevas restricciones en subgrupos de parámetros. Esto produce modelos anidados que se evalúan nuevamente utilizando el test de diferencias en χ^2 . En cada etapa del análisis de invarianza, el test de diferencias de χ^2 compara un modelo restringido, que asume la invarianza en los parámetros que se están evaluando, y un modelo no restringido, el cual no asume invarianza de estos parámetros. La invarianza de los parámetros se acepta si la diferencia en χ^2 no es estadísticamente significativa ($\Delta\chi^2 = \chi^2_{restr} - \chi^2_{norestr}$) a un nivel de significación pre-establecido (por ejemplo, .05).

Dado que el test de χ^2 es sensible al tamaño de la muestra, los investigadores sugieren la utilización de otros estadísticos en la evaluación del cambio en el ajuste entre los distintos niveles de invarianza (Cheung y Rensvold, 2002; Little, 1997; Vandenberg y Lance, 2000). Uno de los índices de ajuste que ha mostrado mayor estabilidad (por ejemplo, en cuanto a tamaños de la muestra y nivel de invarianza), es la diferencia en *CFI* ($\Delta_{CFI} = CFI_{restr} - CFI_{norestr}$). Bajo la asunción de normalidad, Cheung y Rensvold (2002) sugieren que la hipótesis de invarianza no debe mantenerse cuando la diferencia es mayor que .01. Aunque los resultados de Δ_{CFI} no deben ser concluyentes, los investigadores sugieren la utilización complementaria de Δ_{CFI} y $\Delta\chi^2$ para la evaluación de la invarianza de los parámetros entre los grupos.

Algunos índices de ajuste absolutos, como *RMSEA* o *SRMR* han ganado popularidad entre los investigadores que utilizan el AFCM. Chen (2007) ha demostrado que estos dos índices poseen un comportamiento diferencial. Mientras Δ_{RMSEA} ha mostrado ser bastante estable en todos los niveles de invarianza (sugiere una diferencia de .01 como criterio para todos los niveles), Δ_{SRMR} ha mostrado un comportamiento diferente (diferencia de .025 para el *nivel métrico* y .01 para el resto de los niveles de invarianza).

Como hemos visto, aunque existen varios niveles de invarianza, hay consenso general en que el nivel de invarianza métrico es necesario y suficiente para asegurar la E/IM (Meade y Kroustalis, 2006; Vandenberg y Lance, 2000), cuando el interés está focalizado en temas de validez de constructo. Como es bien conocido, las saturaciones factoriales corresponden a las pendientes que relacionan los ítems con

sus correspondientes variables latentes, por lo que representan el cambio esperado en la puntuación observada del ítem por cada unidad de cambio en el factor latente. La existencia de igualdad en las saturaciones factoriales permite establecer la existencia de igualdad en las unidades de cambio entre los ítems y los factores. Esto permite asumir que la métrica de los factores sea equivalente entre los grupos (Jöreskog, 1969; Schmitt, 1982; Vandenberg y Self, 1993). Es por este motivo que la invarianza configural y la invarianza métrica son los dos pasos necesarios para la evaluación de cualquiera de los restantes niveles de invarianza entre los grupos, y condición suficiente para asegurar la E/IM.

En los últimos años la utilización del AFCM se ha visto incrementada en una gran variedad de disciplinas y áreas de investigación (Schmitt y Kuljanin, 2008; Vandenberg y Lance, 2000). Esto se debe principalmente al incremento del número de investigaciones, manuales y software disponibles (Meade y Lautenshlager, 2004). A pesar de ello, los investigadores que realizan comparaciones entre grupos rara vez contrastan la existencia E/IM a través de AFCM en los grupos de análisis. La práctica habitual es basar la validez de las comparaciones sólo en el análisis de aspectos estructurales en la muestra total o en análisis de E/IM realizada en otros grupos (Dimitrov, 2010).

Este aspecto se torna preocupante, especialmente frente a los hallazgos respecto al funcionamiento diferencial de medidas entre distintos grupos. Por ejemplo, Byrne y Cambell (1999) demostraron que un adecuado ajuste de un modelo AFC en la muestra total no garantiza la existencia de E/IM entre los distintos subgrupos que se comparan dentro de dicha muestra. Estos autores observaron que

grupos que mostraron idénticos patrones factoriales, presentaban grandes diferencias en las saturaciones factoriales, en las relaciones entre los factores, en las distribuciones de las puntuaciones, y en los patrones de respuestas de los ítems. Byrne y Cambell (1999) concluyen que la comparación de distintos grupos debe implicar mucho más que el simple ajuste de un modelo AFC y la comparación de las medias.

De este modo, la demostración de la E/IM es un pre-requisito lógico y básico para la evaluación de hipótesis sustantivas con respecto a las diferencias entre grupos, tanto si son complejas contrastes de estructuras teoréticas, como si son simples comparaciones de medias.

A pesar de las posibilidades que brinda el AFCM para la evaluación de la E/IM, la existencia de un efecto de la pertenencia grupal, no necesariamente debería llevar a concluir ausencia de E/IM. En ocasiones es posible que las diferencias encontradas a través de un AFCM no se deban a las diferencias sustantivas en el constructo subyacente, sino al efecto de otros factores que no están relacionados con la estructura interna de los datos. Dentro de estos factores podemos encontrar el efecto del *sesgo de respuesta*.

Stark, Chernyshenko, Chan, Lee y Drasgow (2001) demostraron que una determinada situación de evaluación puede llevar a ciertos grupos al falseamiento de sus respuestas frente al instrumento de medida. La presencia de esta tendencia en un conjunto de datos es conocido como *sesgo de respuesta*. La presencia del *sesgo de respuesta* ha mostrado tener importantes repercusiones en los análisis de fiabilidad y validez de los instrumentos de medida (Cheung y Rensvold, 1999; Marquis, Marquis

y Polich, 1986), y junto con ello, afectar los resultados de los análisis de E/IM. Es por ello que el desarrollo de métodos para detectar, evitar y/o eliminar el efecto del *sesgo de respuesta* ha acaparado el interés de no pocos investigadores.

En el siguiente apartado abordaremos algunas definiciones, tanto desde el punto de vista conceptual como del punto de vista formal, de uno de los tipos de *sesgo de respuesta* más importantes; *el sesgo de respuesta uniforme*. A continuación se presentan también, uno de los principales esfuerzos por eliminar el efecto de este tipo sesgo de respuesta; *la transformación ipsativa o ipsatización* de los datos normativos.

1.3. EL PROBLEMA DEL SESGO DE RESPUESTA

El sesgo de respuesta es una tendencia sistemática a responder parte o la totalidad de un instrumento independientemente tanto de su contenido (Paulhus, 1991) como de la puntuación verdadera del individuo (Fisher, 2004).

En la literatura podemos encontrar varios tipos de sesgo (Cheung y Chan, 2002), pero en general, es posible clasificarlos en dos grandes categorías: el estilo de respuesta (*response style*) y el conjunto de respuesta (*response set*) (Paulhus, 1984; 1991). En el sesgo de *estilo de respuesta*, las respuestas del sujeto pueden verse influenciadas por efectos del contexto, tales como el formato de los ítems o características de los ítems previos. Este fenómeno es llamado frecuentemente *aquiescencia* o *estilo de respuesta aquiescente* (ARS en su acrónimo en inglés). Por otro lado, el sesgo de *conjunto de respuesta* es una reacción temporal a una demanda situacional, por ejemplo, las expectativas sociales. Este estilo de respuesta es llamado

estilo de respuesta extrema (ERS en su acrónimo en inglés). La *deseabilidad social* es un ejemplo de este tipo de *sesgo respuesta*.

En general, se han propuesto dos aproximaciones para tratar de evitar el efecto de estos tipos de sesgo respuesta en el estudio de la estructura interna de un conjunto de datos. Una de ellas se focaliza en el diseño de los ítems, presentando respuestas opuestas o incluyendo alternativas con grado de deseabilidad similar, como es el caso del test de *respuesta forzada*. La segunda aproximación es el control estadístico del sesgo. La estrategia más usada en este contexto es la estandarización del conjunto de datos (Fisher, 2004). El propósito de esta estrategia es el de reducir las diferencias entre los grupos que no sean debidas al constructo de interés, sino producto de la instrumentación (Van de Vijver y Leung, 1997).

Para definir el *sesgo de respuesta*, consideraremos tanto el ARS como el ERS como sesgos de respuesta extremos (Cheung y Chan, 2002). Estudios empíricos muestran que este tipo de sesgos son bastante estables durante largos periodos de tiempo (Greenleaf, 1992), y a través de distintas medidas (Grimm y Church, 1999), lo cual sugiere que la magnitud del sesgo es la misma, o similar, en cada individuo a través de distintas medidas. Este fenómeno se denomina *sesgo de respuesta uniforme* (URB en su acrónimo en inglés).

Formalmente, sean y^* una variable observada e y una variable no observada, con y sin sesgo, respectivamente, entonces (Marquis et al. 1986):

$$y^* = y + 1b, \quad [1.9]$$

donde \mathbf{y}^* e \mathbf{y} son vectores columna de orden $p \times 1$, $\mathbf{1}$ es un vector unitario, y b es una variable aleatoria que corresponde a URB. Ya que \mathbf{y} es una variable no observada, la magnitud de b es desconocida. Cuando b toma valores distintos de 0, entonces las respuestas del entrevistado poseen URB. Además se asume que el URB es independiente del error aleatorio, ya que este tipo de sesgo es independiente del contenido de los ítems (Paulhus, 1991). Este modelo matemático es aplicable tanto al ARS, como al ERS, puesto que la dirección del sesgo podría ser consistente con la dirección de las saturaciones de la estructura factorial (Cheung y Chan, 2002).

Existen varios tipos de estandarización dependiendo de si la transformación involucra la media, la desviación típica, o ambas, o si se realiza con respecto a los individuos o a los grupos (Fisher, 2004). Una de las estrategias sugeridas para reducir el URS es la *transformación ipsativa* o *ipsatización* (Chan y Bentler, 1993; Cheung y Chan, 2002). Este tipo de estandarización corresponde a una estandarización intra-sujeto, ya que ajusta las puntuaciones de cada individuo en función de la media de todas las variables (Hicks, 1970).

Chan y Bentler (1993) proponen dos tipos de transformación ipsativa: la *transformación ipsativa aditiva* y la *transformación ipsativa multiplicativa*. La primera de ellas, que es la de nuestro interés, se obtiene a través de la siguiente expresión:

$$\mathbf{x} = \mathbf{y} - \mathbf{1}\bar{y}, \quad [1.10]$$

donde \mathbf{y} corresponde al vector de datos *normativos* o *pre-ipsativos*, e \bar{y} es el promedio de \mathbf{y} en cada sujeto. Por lo tanto, el vector de datos ipsativos \mathbf{x} es la

desviación de \mathbf{y} desde la media de cada sujeto. Bajo la definición de la ecuación [1.10], el vector \mathbf{x} se denomina vector de *datos ipsativo aditivo* (AID en su acrónimo en inglés).

Matemáticamente, es posible demostrar que la transformación ipsativa aditiva podría eliminar el efecto del sesgo de respuesta uniforme. De acuerdo a Cheung y Chan (2002):

$$\begin{aligned}\mathbf{x}^* &= \mathbf{y}^* - 1\bar{\mathbf{y}}^* \\ &= (\mathbf{y} + 1b) - 1(\bar{\mathbf{y}} + b) \\ &= \mathbf{y} - 1\bar{\mathbf{y}} \\ &= \mathbf{x},\end{aligned}\tag{1.11}$$

donde \mathbf{y}^* e \mathbf{y} corresponden al vector de orden $p \times 1$ con y sin URB, respectivamente (ver ecuación [1.9]), \mathbf{x}^* y \mathbf{x} corresponden a los vectores de datos ipsativos de \mathbf{y}^* e \mathbf{y} respectivamente, e $\bar{\mathbf{y}}^* = \bar{\mathbf{y}} + b$ es el promedio de \mathbf{y}^* . Según esto, es posible observar como la *ipsatización* podría eliminar el efecto de URB del conjunto de datos.

1.4. EL ANALISIS FACTORIAL DE DATOS IPSATIZADOS

1.4.1. El problema del análisis de datos *ipsativos*

A pesar de la posible ventaja de eliminar el URB, los datos *ipsatizados* (y los *ipsativos* en general) tienen algunas propiedades indeseables desde el punto de vista estadístico. Clemans (1966) llama *ipsativo* a cualquier conjunto de datos en los que la suma de las puntuaciones para cada observación individual es igual a una constante. Formalmente:

$$\sum_{i=1}^p x_i = \mathbf{1}' \mathbf{x} = c, \quad [1.12]$$

donde $\mathbf{1}$ es un vector unitario $p \times 1$, y c es una constante. Por tanto, \mathbf{x} es un vector de p dimensiones con propiedades ipsativas. Esta propiedad es conocida como *restricción de suma constante*. A partir de la ecuación [1.12] esta propiedad puede quedar expresada como:

$$\mathbf{1}' \mathbf{x} = \mathbf{1}' (\mathbf{y} - \mathbf{1} \bar{y}) = \mathbf{1}' \mathbf{y} - \mathbf{1}' \mathbf{1} \bar{y} = \mathbf{1}' \mathbf{y} - p \bar{y} = 0 \quad [1.13]$$

Dada esta restricción, los datos ipsativos poseen algunas propiedades estadísticas no deseables que normalmente no se encuentran en un conjunto típico de datos normativos (Chan y Bentler, 1993). Por ejemplo, carecen de una matriz de covarianzas interpretable (Aitchinson, 1986). Dado que $x_1 + x_2 + \dots + x_p = c$, para todo $i = 1, 2, \dots, p$, se cumple que:

$$\begin{aligned} cov(x_i, x_1 + x_2 + \dots + x_p) &= cov(x_i, c) = 0 \\ cov(x_i, x_1) + cov(x_i, x_2) + \dots + cov(x_i, x_i) + \dots + cov(x_i, x_{p-1}) + cov(x_i, x_p) &= 0 \end{aligned} \quad [1.14]$$

A partir de la ecuación [1.14] se deduce que al menos una de las $p-1$ covarianzas debe ser negativa porque $var(x_i)$ es siempre positiva. Esto ocurre en cada fila y columna de la matriz de covarianzas. Por otro lado, se sabe que la suma de las varianzas y covarianzas es cero en todas las filas y columnas de Σ_x . Por tanto, si Σ_x es la matriz de covarianzas de \mathbf{x} , entonces:

$$\Sigma_x \mathbf{1} = 0, \text{ ó } \mathbf{1}' \Sigma_x = 0'. \quad [1.15]$$

Esta propiedad tiene importantes implicaciones para el estudio de la estructura de covarianzas de un conjunto de datos. Se sabe que la suma de los términos de una matriz de covarianzas de datos *ipsativos* es cero en cada una de sus filas y columnas. Esto conduce a dos importantes consecuencias. Primero, los datos ipsativos generan estructuras bipolares artificiales debido a la interdependencia de las variables, y los métodos correlacionales estándar fallan en reflejar la estructura verdadera de la población (Baron, 1996; Closs, 1996; Cornwell y Dunlap, 1994; Dunlap y Cornwell, 1994; Loo, 1999). Y segundo, y más importante, en el área del análisis de la estructura de covarianzas, no es posible la aplicación de las técnicas multivariantes clásicas para este tipo de datos, las cuales han sido desarrolladas para vectores sin este tipo de restricción, o al menos no directamente, ya que pueden llevar a interpretaciones erróneas y a inferencias dudosas (Aitchinson, 1986). Más específicamente, en el área del análisis de la estructura de covarianzas, los procedimientos de estimación no pueden llevar a cabo los análisis si la matriz de covarianzas de entrada es singular. Por ejemplo, para el caso de GLS (estimación por Mínimos Cuadrados Generalizados), la función $F_{GLS} = 1 / 2tr[(\mathbf{S} - \mathbf{\Sigma})\mathbf{S}^{-1}]$, requiere que \mathbf{S} sea regular. Mientras que para ML (Máxima Verosimilitud), el cálculo del estadístico de ajuste χ^2 depende del determinante de \mathbf{S} (Chan y Bentler, 1996).

1.4.2. El AFC de datos *ipsatizados*

En un intento de solución, Jackson y Alwin (1980) abordaron el problema del análisis de los datos AID en el contexto del análisis factorial exploratorio. Asumiendo

la existencia de un vector subyacente de datos *normativos* (o *pre-ipsativos*) y que siguen un modelo de análisis factorial con parámetros θ , propusieron una posible solución para estimar los parámetros basado en el análisis de x . Partiendo de la asunción de que x es un vector $p \times 1$ de los AID, e y es su correspondiente vector de datos *normativos* o *pre-ipsativos*, y siguiendo la transformación ipsativa definida en la ecuación [1.10];

$$\begin{aligned} x &= y - 1\bar{y} \\ &= y - 1p^{-1}1'y \\ &= y - p^{-1}11'y \\ &= (I - p^{-1}11')y \\ x &= Ay, \end{aligned} \tag{1.16}$$

donde

$$A = I - p^{-1}11', \tag{1.17}$$

I es una matriz de identidad de orden p , y como la dimensión de p es conocida A está completamente determinada. Desde la ecuación [1.16] se puede notar que x e y están estrechamente relacionadas a través de la matriz A .

Siguiendo la formulación del Análisis Factorial Confirmatorio Convencional (Jöreskog y Sörbom, 1981):

$$y = \mu_y + \Lambda_y \xi + \varepsilon, \tag{1.18}$$

donde y es un vector de p variables observadas, ξ es un vector q factores, μ_y es $E(y)$, Λ_y es la matriz de saturaciones factoriales de orden $p \times q$, y ε es el vector de

p errores de medida. Aquí se asume que $E(\xi)=0$, $E(\varepsilon)=0$ y $E(\xi\varepsilon')=0$. Además la matriz de covarianzas de y es:

$$\begin{aligned}\Sigma_y &= E\left((y - \mu_y)(y - \mu_y)'\right) \\ &= \Lambda_y \Phi_y \Lambda_y' + \Psi_y ,\end{aligned}\quad [1.19]$$

donde Φ_y corresponde a la matriz de covarianzas de los q factores y Ψ_y a la matriz de covarianzas de los p de errores de medida. Para recuperar la estructura normativa de x , Chan y Bentler (1993) sugieren la elaboración de una estructura de vinculación entre x y el modelo AFC de la ecuación [1.18] referido a y .

A partir de la expresión $x = Ay$ definida en la ecuación [1.16], e incluyendo la ecuación [1.18], el modelo AFC para x puede expresarse mediante:

$$\begin{aligned}x &= Ay \\ x &= A(\mu_y + \Lambda_y \xi + \varepsilon) \\ &= \mu_x + A\Lambda_y \xi + A\varepsilon \\ &= \mu_x + (I - p^{-1}11')\Lambda_y \xi + A\varepsilon \\ &= \mu_x + (\Lambda_y - 1\bar{\lambda}')\xi + A\varepsilon \\ &= \mu_x + \Lambda_x \xi + \delta ,\end{aligned}\quad [1.20]$$

donde

$$\mu_x = E(x) = A\mu_y \quad [1.21]$$

y

$$\Lambda_x = A\Lambda_y = \Lambda_y - 1\bar{\lambda}' , \quad [1.22]$$

siendo $\overline{\lambda'}$ un vector fila de promedios de los coeficientes de cada columna de la matriz Λ_y .

De forma similar, siguiendo la ecuación $x = Ay$, la matriz de covarianzas de x puede expresarse como:

$$\begin{aligned}\Sigma_x &= A\Sigma_y A' \\ &= A(\Lambda_y \Phi_y \Lambda_y' + \Psi_y)A \\ &= \Lambda_x \Phi_y \Lambda_x' + \Psi_\varepsilon\end{aligned}\tag{1.23}$$

A través de las ecuaciones [1.20] y [1.23], Jackson y Alwin argumentaron que el modelo de y está bien definido, y el modelo de x también puede estar identificado. Por otro lado, a través de la ecuación [1.23] podemos ver que la matriz de covarianzas de los factores Φ_y , se mantiene sin variación, y que los errores de medida se mantendrán relacionados con el modelo ipsativo a través de A .

Nótese que debido a las propiedades ipsativas de x , la matriz Σ_x es singular (ver ecuación [1.15]), y por tanto, y no puede ser analizada directamente. Para solucionar este problema Jackson y Alwin sugieren que se elimine arbitrariamente una de las columnas (y su correspondiente fila) de Σ_x para así obtener una matriz no singular Σ_x^* , la cual es una submatriz de Σ_x . Esto puede llevarse a cabo a través de:

$$\Sigma_x^* = K\Sigma_x K,\tag{1.24}$$

donde K es una matriz de selección $(p-1) \times p$. Como resultado, tanto Φ_y como Ψ_ε pueden estimarse satisfactoriamente ya que ambas están incluidas en la ecuación [1.23]. Sin embargo, las saturaciones factoriales Λ_y no pueden estimarse debido a

que la matriz de saturaciones factoriales Λ_x ahora contiene las desviaciones desde el promedio de las saturaciones factoriales de cada columna de la matriz Λ_y (ver ecuación [1.22]). Ello puede llevar a problemas de interpretación del modelo original ya que tras la substracción de $\overline{1\lambda'}$ desde Λ_y , el signo original de las saturaciones factoriales puede quedar alterado. Para resolver el problema, Chan y Bentler (1993) proponen una solución en el contexto del AFC. Ellos sugieren la creación de una estructura de vinculación entre x y el modelo normativo AFC de y , siendo los parámetros del modelo desconocidos.

Al comparar el modelo normativo de la ecuación [1.18] con el modelo ipsativo de la ecuación [1.20], es posible reparametrizar el modelo *ipsativo* utilizando las ecuaciones [1.20], [1.21] y [1.22]. De este modo, podemos recuperar los parámetros del modelo normativo. Sin embargo la re-parametrización requiere de cálculos externos para su obtención (para más detalles sobre el cálculo manual para la recuperación de los parámetros del modelo normativo, así como el cálculo de los errores estándar, el lector puede consultar a Cheung, 2004, pág. 238).

Este método requiere tediosos cálculos para recuperar los parámetros del modelo y sus errores estándar, además de no permitir la introducción de restricciones inter-grupo. Para solucionar estos problemas, Cheung y Chan (2004) desarrollaron un método de estimación directa (en adelante, el método *DE*), que es matemáticamente equivalente al método de Chan y Bentler (1993).

El método *DE* propone un modelo de análisis factorial de 2° orden:

$$\begin{aligned}\boldsymbol{\eta} &= \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \text{ y} \\ \mathbf{x} &= \mathbf{A}\boldsymbol{\eta},\end{aligned}\tag{1.25}$$

donde $\boldsymbol{\eta}$ es el vector de los p factores de 1^{er} orden, $\boldsymbol{\xi}$ es el vector de los q factores de 2° orden, $\boldsymbol{\Gamma}$ es la matriz de saturaciones factoriales de 2° orden y $\boldsymbol{\zeta}$ es un vector de los p errores asociados a $\boldsymbol{\eta}$. De este modo, $\boldsymbol{\Gamma}$ contiene los parámetros estructurales que corresponden a las saturaciones factoriales de 1^{er} orden del modelo AFC normativo, y se obtienen directamente, sin necesidad de cálculos externos para recuperarlos. El modelo factorial de 1^{er} orden ($\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}$) fija los parámetros de $\boldsymbol{\Lambda}_y$ a los valores de la matriz \mathbf{A} y estima libremente los de $\boldsymbol{\varepsilon}$.

1.4.3. Implementación del método *DE* en un modelo SEM

Para ilustrar la implementación del método *DE* utilizaremos el ejemplo de un modelo dos factores de tres variables por factor, por ser un modelo sencillo. Las ecuaciones del modelo estructural de la ecuación [1.25]) quedan definidas para el ejemplo del siguiente modo:

$$\begin{aligned}\eta_1 &= \gamma_1\xi_1 + \zeta_1 \\ \eta_2 &= \gamma_2\xi_1 + \zeta_2 \\ \eta_3 &= \gamma_3\xi_1 + \zeta_3 \\ \eta_4 &= \gamma_4\xi_2 + \zeta_4 \\ \eta_5 &= \gamma_5\xi_2 + \zeta_5, \text{ y} \\ \eta_6 &= \gamma_6\xi_2 + \zeta_6.\end{aligned}\tag{1.26}$$

Puesto que se necesita eliminar una variable de la matriz de covarianzas para evitar la singularidad, sólo utilizaremos cinco variables observadas. Según esto, las ecuaciones del modelo de medida quedan como:

$$\begin{aligned} x_1 &= .883\eta_1 - .167\eta_2 - .167\eta_3 - .167\eta_4 - .167\eta_5 - .167\eta_6 \\ x_2 &= -.167\eta_1 + .883\eta_2 - .167\eta_3 - .167\eta_4 - .167\eta_5 - .167\eta_6 \\ x_3 &= -.167\eta_1 - .167\eta_2 + .883\eta_3 - .167\eta_4 - .167\eta_5 - .167\eta_6 \\ x_4 &= -.167\eta_1 - .167\eta_2 - .167\eta_3 + .883\eta_4 - .167\eta_5 - .167\eta_6, \text{ y} \\ x_5 &= -.167\eta_1 - .167\eta_2 - .167\eta_3 - .167\eta_4 + .883\eta_5 - .167\eta_6, \end{aligned} \quad [1.27]$$

las cuales corresponden al sistema que vincula las variables observadas \mathbf{x} con las variables latentes de 1^{er} orden $\boldsymbol{\eta}$ (matriz \mathbf{A}).

El diagrama *path* del modelo AFC de 2^o orden se muestra en la figura 1. Si observamos el modelo *ipsativo* y el método *DE*, podemos deducir algunas propiedades importantes. En primer lugar, las vinculaciones entre $\boldsymbol{\xi}$ y $\boldsymbol{\eta}$ en el método *DE* corresponden a las saturaciones factoriales del modelo *preipsativo*. En segundo lugar, los errores de $\boldsymbol{\xi}$ hacia $\boldsymbol{\eta}$ en el método *DE* corresponden directamente a los errores de medida del modelo *preipsativo*. Y en tercero, las propiedades *ipsativas* de los datos se reflejan en las saturaciones factoriales fijadas desde $\boldsymbol{\eta}$ hasta \mathbf{x} (ver ecuación [1.27]).

Como se ha mencionado anteriormente, aunque este procedimiento es matemáticamente equivalente al método de Chan y Bentler (1993), el método *DE* (Cheung, 2004) ofrece importantes ventajas, no sólo en la estimación directa de los parámetros y los errores de estimación. En el contexto del análisis E/IM, el método *DE* permite la imposición de restricciones de igualdad inter-grupo, lo cual permite la aplicación del AFCM al análisis de variables *ipsatizadas*.

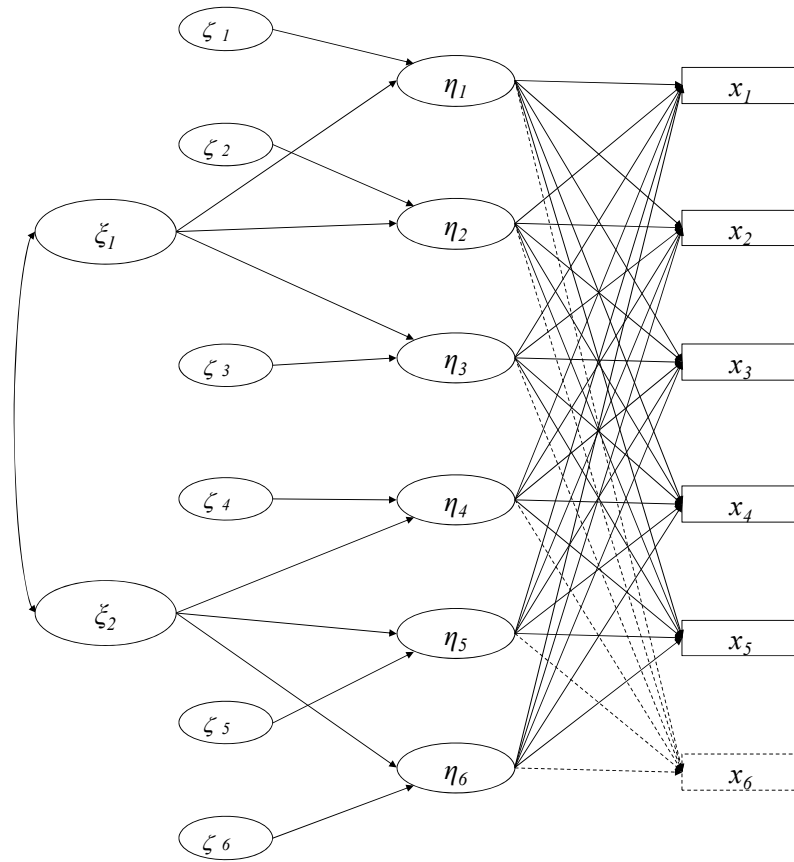


Figura 1. Modelo factorial del método DE de Cheung (2004) para el modelo ipsativo de 2 factores. La variable x_6 y las líneas discontinuas no son incluidas en el análisis con el fin de evitar la singularidad de la matriz de covarianza de los datos ipsativos. Tomado de Cheung (2004, pag. 228).

Junto con presentar la formulación analítica del método *DE* e ilustrar su aplicación en un grupo único, Cheung (2004) presenta la extensión de su propuesta para el caso del análisis multi-grupo e ilustra su implementación en un estudio con datos reales. Dado que la conformación de los grupos para su ejemplo se llevó a cabo de manera aleatoria, los resultados del ajuste del modelo fueron razonables. Por tanto, Cheung (2004) concluye que el método *DE* y su extensión al AFCM, posee un funcionamiento satisfactorio.

A pesar del adecuado funcionamiento en la estimación de parámetros y en la detección de la invarianza entre grupos, hasta la fecha no existen estudios que evalúen la efectividad del análisis mediante la *ipsatización* para eliminar el *sesgo de respuesta* en el AFCM. Tal como sugiere Cheung (2004), se necesita investigación futura que intente evaluar el rendimiento de AFCM en el análisis de variables *ipsatizadas*.

CAPITULO II

LOS TEST IPSATIVOS DE RESPUESTA FORZADA

En este capítulo se presenta una segunda propuesta para evitar el *sesgo de respuesta* de un conjunto de datos: el formato de *respuesta forzada*. En primer lugar, describiremos sus principales características, así como la naturaleza analítica de los datos que produce. A continuación, resumiremos una de las actuales propuestas para el análisis de esta clase de test. Se ha dedicado un capítulo a este tipo de formato de respuesta porque lo utilizaremos posteriormente en nuestro estudio empírico para comparar sus resultados con los resultados del AFCM de datos *ipsatizados*. Para ello utilizaremos dos versiones de un mismo test, una *normativa* y otra de *respuesta forzada*.

Otra de las estrategias antes mencionadas para tratar el *sesgo de respuesta* es la aproximación instrumental. Uno de los esfuerzos más prometedores, corresponde a los test de formato de *respuesta forzada*. En este tipo de test, el entrevistado debe ordenar alternativas de grado de deseabilidad similar. Con ello se intenta forzar a que la persona deba priorizar cada una de las alternativas en función a las demás.

Uno de los test de *respuesta forzada* más conocidos es el “Inventario de Estilos de Aprendizaje”, LSI de Kolb (1985), un test ampliamente utilizado en contextos educativos. Los entrevistados deben ordenar jerárquicamente cuatro alternativas de respuesta presentadas en cada una de las preguntas o bloques. Este test

consta de 12 preguntas o bloques de tipo auto-informe que deben ordenarse en cuatro opciones de respuesta (4, 3, 2 y 1), desde la que “*más me representa*” (4) a la que “*menos me representa*” (1), respondiendo a la pregunta “¿*Cómo aprendo mejor?*”. Cada alternativa del bloque corresponde a uno de los cuatro estilos de aprendizaje que mide el test: *acomodador*, *divergente*, *asimilador* y *convergente*. Como resultado, tanto la puntuación total de cada bloque ($1 + 2 + 3 + 4 = 10$), como la puntuación total del test ($10 \times 12 = 120$), es la misma para todos los sujetos.

El formato de respuesta de esta clase de test involucra un ordenamiento (*ranking*) de los ítems, a través de un número fijo de opciones de respuesta en cada bloque. La forma tradicional de codificar las respuestas, utilizando las mismas puntuaciones en cada uno de los bloques, conduce a que todos los entrevistados reciban la misma puntuación en cada bloque, y junto con ello, la misma puntuación total en el test. Es por ello que los test de *respuesta forzada* producen datos con propiedades *ipsativas*. A diferencia de la *ipsatización*, en donde los datos son obtenidos en forma *normativa* para luego ser *ipsatizados*, en el formato de *respuesta forzada* los datos son obtenidos directamente de forma *ipsativa*.

A continuación se exponen las principales características de los test de *respuesta forzada*, y cómo este tipo de test genera datos con propiedades *ipsativas*. Junto con ello, se presenta el modelo analítico, basado en la *Ley del Juicio Comparativo* de Thurstone (1931), que proporciona un marco conceptual de gran utilidad para comprender la naturaleza de este tipo de datos. Finalmente, se presenta la formulación analítica de uno de los intentos más recientes de analizar la estructura interna de los datos *ipsativos* producto del formato de *respuesta forzada*.

2.1. EL FORMATO DE RESPUESTA FORZADA

A diferencia del formato Likert, en el que se responde a un enunciado seleccionando una alternativa de un conjunto de categorías ordenadas, en el formato de *respuesta forzada* el sujeto debe ordenar distintos enunciados de acuerdo a cómo describen su comportamiento, respondiendo a todas las alternativas del bloque.

La figura 2 muestra uno de los bloques del test mencionado en el ejemplo anterior. Cada alternativa de respuesta mide una de las cuatro dimensiones medidas por el test. El sujeto debe ordenar las alternativas desde “la que más le representa” (4) hasta “la que menos le representa” (1). Cada bloque (p) está compuesto por $n = 4$ ítems que miden $d = 4$ dimensiones. Por ejemplo, con $p = 12$ bloques de $n = 4$ alternativas por bloque, el test tendría $m = 48$ ítems en total ($n \times p = 48$).

En lugar de evaluar cada enunciado por separado, el entrevistado debe responder a cada alternativa en relación a las demás alternativas del bloque, lo cual produce una interrelación de las puntuaciones entre los ítems que pertenecen al mismo bloque. Puesto que no es posible asignar a las alternativas del mismo bloque la misma puntuación, el formato de respuesta forzada contribuye a eliminar el sesgo de respuesta producto de la *acquiescencia* (Cheung y Chan, 2002) y a reducir el *efecto halo* (Bartram, 2007), pues el entrevistado debe priorizar una alternativa sobre las restantes. Sin embargo, los datos resultantes tienen propiedades *ipsativas*.

Yo aprendo mejor...			
<input type="checkbox"/> ...sintiendo	<input type="checkbox"/> ...observando	<input type="checkbox"/> ...pensando	<input type="checkbox"/> ...haciendo

Figura 2. Ejemplo de un ítem de respuesta forzada. Adaptado de Kolb(1985)

2.2. EL MODELO THURSTONIANO

El modelo de Thurstone (1931) proporciona un marco conceptual que permite comprender la naturaleza de los datos *ipsativos de respuesta forzada*. Según la *Ley del Juicio Comparativo* de Thurstone, frente a la tarea de ordenar distintos objetos (*items*), el sujeto debe realizar un proceso de comparación. El modelo propone tres supuestos básicos: 1) cada objeto que se compara (*item*), provoca una variable latente (*utilidad*) como resultado del proceso de discriminación; 2) Al comparar dos objetos, el sujeto elige el objeto con mayor nivel de *utilidad*, y 3) la *utilidad* corresponde a una variable latente con distribución normal.

Sea t_i la *utilidad* latente asociada al ítem i , el sujeto preferirá el ítem i sobre otro ítem k , si la *utilidad* latente del primero es mayor que la del segundo. Según esto, la preferencia se puede codificar como:

$$y_l = \begin{cases} 1 & \text{si } t_i \geq t_k \\ 0 & \text{si } t_i < t_k \end{cases}, \quad [2.1]$$

donde y_l corresponde al resultado de la comparación por pares de los ítems i y k .

Este modelo puede extenderse fácilmente a la tarea de ordenamiento de más de dos objetos. Frente a p alternativas de respuesta, el sujeto debe ordenar los ítems desde p (el que más le representa) hasta 1 (el que menos le representa). El modelo Thurstoniano asume la existencia de un vector $Y = (Y_1, \dots, Y_p)'$ de p variables aleatorias, que posee distribución normal multivariante. De este modo el sujeto ordena los objetos según su nivel en la variable Y_i .

Este modelo permite asumir la existencia de una variable con propiedades normativas (*utilidades*) que subyace a cada objeto (*ítem*) en la tarea de ordenamiento.

Por tanto, cada sujeto utiliza el nivel de *utilidad* que subyace a cada objeto para realizar la tarea de ordenamiento en un bloque de *respuesta forzada*. De este modo, se asume un modelo factorial convencional de variables con propiedades *normativas*, relacionado con los datos *ipsativos* producto del formato de *respuesta forzada*.

2.3. ANÁLISIS FACTORIAL DE ÍTEMS DE RESPUESTA FORZADA

Uno de los esfuerzos más prometedores para resolver los problemas analíticos en el análisis de la estructura interna de los datos de *respuesta forzada* es el de Maydeu-Olivares y Böckenholt (2005). Los autores sostienen que el problema radica en la codificación tradicional de los datos obtenidos a través de este tipo de formato. Su propuesta se basa en un procedimiento de codificación binaria. En un bloque de *respuesta forzada*, el sujeto debe ordenar los n ítems de acuerdo a sus preferencias. Siguiendo con el ejemplo de un bloque de $n = 4$ ítems (A, B, C y D), el sujeto debe asignar la posición de cada ítem en relación a sus preferencias, tal y como se muestra en la figura 3.

Ordene los cuatro enunciados desde el 4 (“*el que más me representa*”) hasta el 1 (“*el que menos me representa*”)

<input type="checkbox"/>	Ítem A	<input type="checkbox"/>	Ítem B	<input type="checkbox"/>	Ítem C	<input type="checkbox"/>	Ítem D
--------------------------	--------	--------------------------	--------	--------------------------	--------	--------------------------	--------

Figura 3. Formato de un bloque de respuesta forzada de 4 ítems ($n = 4$)

En cualquier bloque de n ítems, es posible codificar las respuestas a través de una codificación binaria usando $\tilde{n} = n(n - 1) / 2$ pares de comparaciones. En nuestro ejemplo, en un bloque de $n = 4$ podemos realizar $\tilde{n} = 4(4 - 1) / 2 = 6$

comparaciones binarias entre los 4 ítems del bloque. En todos los pares de comparaciones podemos utilizar el siguiente formato de codificación:

$$y_l = \begin{cases} 1 & \text{si se prefiere el ítem } i \text{ sobre el ítem } k, \\ 0 & \text{si se prefiere el ítem } k \text{ sobre el ítem } i. \end{cases} \quad [2.2]$$

Por tanto, en nuestro ejemplo la codificación del bloque de $n = 4$ quedaría de la forma presentada en la figura 4.

Respuesta forzada				Codificación binaria					
A	B	C	D	(A,B)	(A,C)	(A,D)	(B,C)	(B,D)	(C,D)
3	1	2	4	1	1	0	0	0	0

Figura 4. Formato de codificación de un bloque de respuesta forzada

Como propone la *Ley del Juicio Comparativo* de Thurstone, cada comparación por pares, genera una *utilidad* para cada objeto que se compara según la ecuación [2.1]. A partir de esta ecuación el proceso de codificación puede describirse de manera alternativa;

$$y_l^* = t_i - t_k, \quad [2.3]$$

donde y_l^* corresponde a una variable continua que representa la diferencia entre las *utilidades* de los ítems presentes en cada comparación. Por tanto, la vinculación entre la respuesta de comparación por pares y_l y la respuesta de comparación latente y_l^* es

$$y_l = \begin{cases} 1 & \text{si } y_l^* \geq 0, \\ 0 & \text{si } y_l^* < 0. \end{cases} \quad [2.4]$$

Expresado en términos matriciales, sea \mathbf{t} un vector de n *utilidades* latentes, e \mathbf{y}^* un vector de \tilde{n} diferencias latentes, donde $\tilde{n} = n(n-1)/2$ corresponde a la cantidad de comparaciones por pares, las \tilde{n} ecuaciones quedan expresadas como:

$$\mathbf{y}^* = \mathbf{A}\mathbf{t}, \quad [2.5]$$

donde \mathbf{A} es una matriz de diseño que vincula las diferencias latentes con las *utilidades* latentes. Cada fila de \mathbf{A} corresponde a cada uno de los \tilde{n} pares de comparaciones y cada columna a cada uno de los n ítems.

Para el caso de un bloque de 2 ítems $n = 2$: $\mathbf{A} = [1 -1]$.

$$\text{Para } n = 3: \mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}, \text{ y para } n = 4: \mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

La matriz de diseño especifica cuáles son los ítems que participan en cada comparación. Para el caso del bloque de 4 ítems ($n = 4$), cada columna corresponde a cada uno de los 4 ítems (A, B, C y D). Por otro lado, cada fila de la matriz representa cada una de las 6 posibles comparaciones de los tres ítems del bloque. La primera fila compara los ítems {A, B}, la segunda fila los ítems {A, C}, la tercera fila compara los ítems {A, D}, la cuarta fila a los ítems {B,C}, la quinta a los ítems {B,D}, y finalmente la sexta fila a los ítems {C,D}. Para el caso de un cuestionario de múltiples bloques, tenemos p número de bloques, n número de ítems por bloque y $n \times p = m$ número de ítems en total. Por tanto, la matriz de diseño tendrá m columnas correspondientes al total de ítems del cuestionario y $p \times \tilde{n}$ filas correspondientes a los \tilde{n} pares de comparaciones realizadas en todos los p bloques. Para el ejemplo presentado anteriormente, el cuestionario de $p = 3$ bloques, de

$n = 4$ ítems por bloque, $m = 12$ ítems en total, la matriz de diseño quedaría del siguiente modo:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

Por otro lado, siguiendo el modelo de AFC clásico, que se expresa mediante (Jöreskog y Sörbom, 1981):

$$\mathbf{y} = \boldsymbol{\mu}_y + \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad [2.6]$$

nuestro modelo de AFC, que mide d factores comunes a través de las *utilidades* latentes \mathbf{t} , quedaría expresado como:

$$\mathbf{t} = \boldsymbol{\mu}_t + \mathbf{A}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad [2.7]$$

donde $\boldsymbol{\mu}_t$ es el vector que contiene las m medias de las *utilidades* latentes \mathbf{t} , \mathbf{A} es la matriz $m \times d$ de coeficientes factoriales, $\boldsymbol{\eta}$ es el vector de factores comunes y $\boldsymbol{\varepsilon}$ es

un vector de los m factores únicos (errores de medida en el AFC convencional). Como en un AFC clásico, se asume que: 1) cada ítem, y por tanto cada *utilidad*, mide solo un factor; 2) los factores no están correlacionados con las unicidades; 3) la media de los factores comunes es cero; 4) la correlación entre los factores (matriz Φ) son parámetros libres a estimar y sus varianzas se fijan a 1 para la identificación del modelo; 4) la matriz Ψ de unicidades es diagonal, lo cual indica que las unicidades no están correlacionadas; y finalmente, 5) tanto los factores comunes como las unicidades presentan distribución normal.

Combinando las ecuaciones [2.5] y [2.7] obtenemos un modelo factorial de 2° orden que vincula los resultados binarios observados con los factores comunes subyacentes.

La figura 5 ilustra cómo se modelan los resultados binarios, las *utilidades* y los factores latentes en el ejemplo. El modelo corresponde a una estructura sencilla de $d=4$ factores correlacionados, los cuales se definen mediante 3 ítems por factor. Los $m=12$ ítems del cuestionario se presentan en bloques de $n=4$ ítems. Nótese que no hay más de un ítem por bloque que mida el mismo factor latente.

El cuestionario está formado por $p=3$ bloques. El primer factor está medido por los ítems 1, 5 y 9, el segundo por los ítems 2, 6 y 10, el tercero por los ítems 3, 7 y 11, y el cuarto por los ítems 4, 8 y 12. La tarea consiste en ordenar los 4 ítems de cada bloque de acuerdo a su preferencia (desde 4, el que más le representa; a 1, el que menos le representa). 18 en total, los cuales se modelan como diferencias de pares de *utilidades* subyacentes a través de la ecuación [2.5].

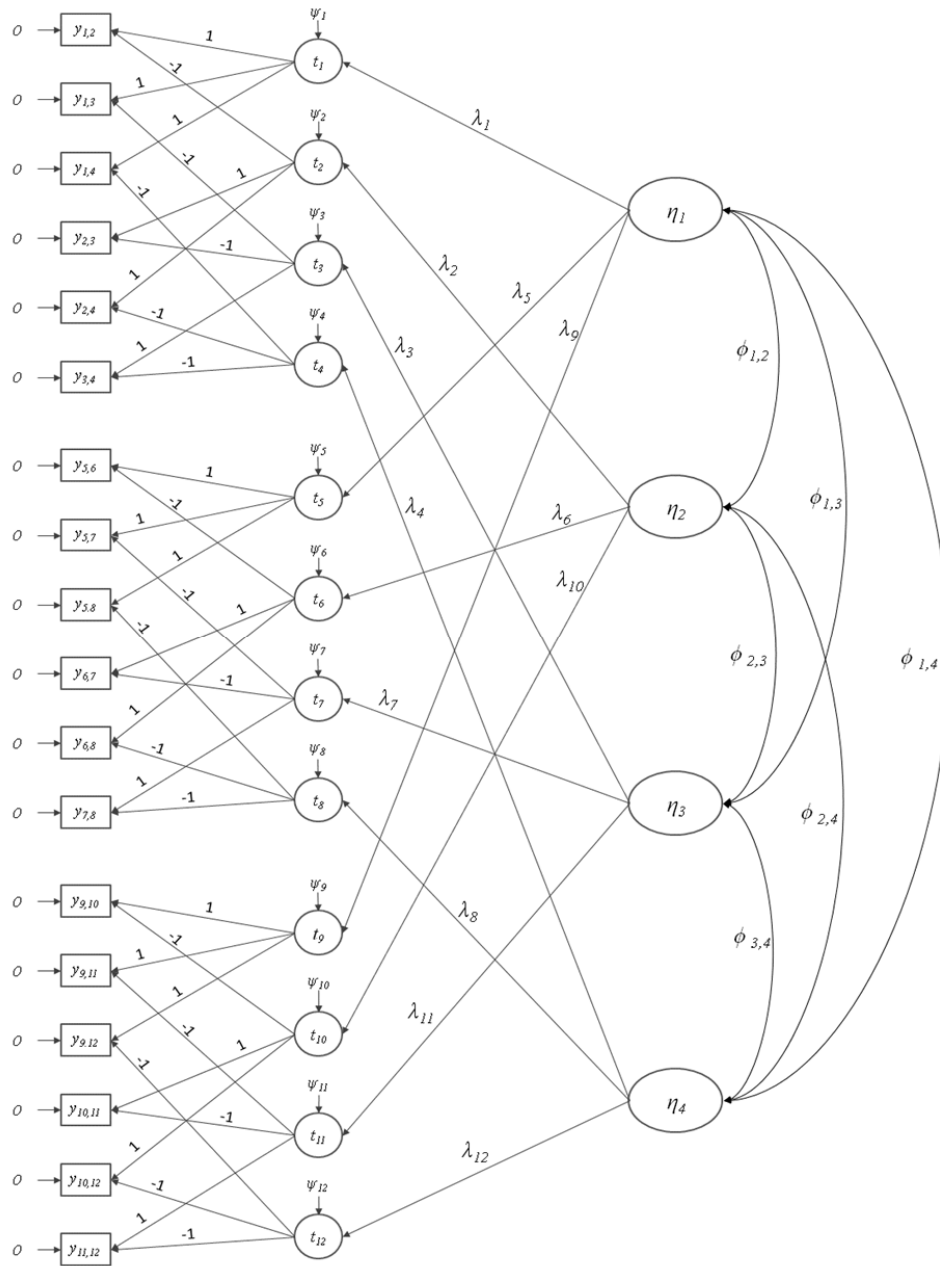


Figura 5. Modelo factorial Thurstoniano para un cuestionario de respuesta forzada de 4 factores con tres bloques de 4 ítems por bloque (basado en Brown y Maydeu-Olivares, 2011, pág. 467).

Como se observa, cada resultado binario ($y_{i,k}$) conlleva la comparación de dos ítems y depende de las dos utilidades latentes de cada ítem. A su vez, cada

utilidad latente es función de un factor latente. Nótese que los 18 resultados binarios están medidos sin error; sin embargo, la varianza de las 12 *utilidades* latentes son parámetros libres a estimar, las cuales corresponden a las unicidades de cada *utilidad* (los errores de medida de un AFC convencional). De este modo, estamos frente a un modelo factorial de segundo orden que se define mediante 18 variables observadas que corresponden a los resultados binarios medidos sin error, 12 factores de primer orden que corresponden a las *utilidades* latentes asociadas a un error de predicción (*unicidad*), y 4 factores de segundo orden que equivaldrían a los factores latentes de primer orden en un AFC convencional.

Los resultados del ordenamiento se transforman a 6 resultados binarios por bloque. Para la identificación del modelo Maydeu-Olivares y Böckenholt (2005) ofrecen algunas sugerencias para ajustar datos provenientes de formatos de respuesta forzada. Para el caso de modelos multidimensionales basta con dos tipos de restricciones: 1) fijar la varianza de los factores latentes a 1, como en un modelo factorial convencional; y 2) puesto que las *utilidades* son variables latentes, hay que fijar la métrica de sus errores; para ello es necesario fijar el error de una *utilidad* a 1 en cada bloque. En este caso, ya que la *utilidad* latente depende de dos resultados binarios, el error de una de las *utilidades* por cada bloque puede ser fijado a .5, para cumplir $\psi_l^2 = \psi_i^2 + \psi_k^2$.

Recientemente Brown y Maydeu-Olivares (2011), siguiendo el trabajo de Maydeu-Olivares y Böckenholt (2005), han adaptado el modelo AFC Thurstoniano a un modelo de TRI multidimensional con algunas características especiales. Aunque

este modelo es matemáticamente equivalente, posee importantes ventajas en relación al modelo Thurstoniano de 1^{er} orden.

En primer lugar, desde el punto de vista psicométrico, el interés no está en la estimación de los factores de primer orden (*utilidades*), sino más bien en las estimaciones de los factores de segundo orden (*factores comunes*). En segundo lugar, el uso de un modelo TRI Thurstoniano, en vez de un modelo Thurstoniano de 2^o orden, agiliza considerablemente los cálculos en el caso de modelos con mayor número de variables. Finalmente, en el modelo Thurstoniano de 2^o orden, debido a que las varianzas de los errores asociados a y^* son iguales a cero, los factores latentes no pueden ser estimados (Maydeu-Olivares, 1999). Cuando el modelo es reparametrizado a un modelo de 2^o orden, las varianzas de los errores de la *utilidades* latentes son distintas de cero, lo que permite la estimaciones de los factores latentes (para detalles sobre el procedimiento para ajustar un modelo TRI Thurstoniano ver Brown y Maydeu-Olivares, 2011).

Brown y Maydeu-Olivares (2011) encontraron que el modelo de TRI multidimensional Thurstoniano permite la adecuada recuperación del modelo, la estimación directa del nivel de rasgo de los individuos y la obtención de la función de información del test en la mayoría de las condiciones que fueron estudiadas. En sus estudios de simulación evaluaron una serie de condiciones relacionadas con el diseño de test de *respuesta forzada*, a saber; dirección de los ítems (todos positivos v/s mezcla de positivos y negativos), tamaños de los bloques (2, 3 y 4 ítems por bloque), correlación entre los factores (-.5; 0 y .5), y número de factores (2, 3, 4 y 5). Los resultados mostraron que la recuperación tanto de los parámetros del modelo, como

del nivel de rasgo de los sujetos mejora a medida que aumenta el número de factores en el modelo ($d \geq 3$), el número ítems en el bloque ($n \geq 3$), la correlación entre los factores es igual o menor que cero ($r \leq 0$) y la dirección de los ítems en un mismo bloque es opuesta. A pesar de ello, el único diseño que no recomiendan los autores debido a los pobres resultados obtenidos es el de bloques de 2 ítems. Finalmente, en relación a la tasa de rechazo de la hipótesis de ajuste según χ^2 , los resultados muestran que en el modelo de más de 1000 grados de libertad, el estadístico χ^2 infra-estima el grado de ajuste del modelo, aún cuando los parámetros de los ítems y sus errores son adecuadamente estimados.

A pesar de que frente a algunas de las condiciones el modelo de TRI multidimensional Thurstoniano muestra un funcionamiento adecuado, las condiciones que han sido estudiadas son aún reducidas, y los hallazgos no pueden ser generalizados a otras condiciones. Por tanto, creemos que es todavía necesaria mucha más evidencia, especialmente en situaciones reales, para asegurar el adecuado funcionamiento de este modelo. Por otro lado, también es necesario el desarrollo de técnicas que solucionen los problemas de estimación para modelos demasiado grandes o de más de mil grados de libertad¹.

¹ Debido al proceso de codificación binaria, los grados de libertad del modelo no sólo dependen de la cantidad de variables y de la cantidad de parámetros libres a estimar, sino también del tamaño de los bloques. Por ejemplo, un bloque de tres ítems se codifica en $\tilde{n} = 3(3-1)/2 = 3$ resultados binarios. Por el contrario, un bloque de cuatro alternativas se codifican en $\tilde{n} = 4(4-1)/2 = 6$ resultados binarios, lo que duplica el tamaño de la matriz de covarianzas.

CAPITULO III

ANTECEDENTES Y ESTADO ACTUAL DEL PROBLEMA

En este capítulo se realiza una revisión de los escasos estudios existentes hasta la fecha con respecto al AFC de variables *ipsatizadas*, poniendo énfasis en los factores que afectan a la sensibilidad de los índices de ajuste en la evaluación de este tipo de modelos. Junto con ello, se presenta una revisión de la investigación previa con respecto a los factores que afectan al funcionamiento del AFCM con datos *normativos*.

La investigación previa acerca de las condiciones óptimas para llevar a cabo un AFC con variables *ipsatizadas* es escasa. Tan sólo existen seis estudios publicados que han examinado la factorización de variables con este tipo de datos (Chan, 2003; Cheung, 2004, 2006; Chan y Bentler, 1993, 1996; Ximénez y Calderón, 2012), y sólo dos de ellos evalúan factores que afectan al funcionamiento del AFC con variables *ipsatizadas*.

3.1. FUNCIONAMIENTO DEL AFC DE VARIABLES IPSATIZADAS

Como ya se expuso en el capítulo anterior, Chan y Bentler (1993) plantearon una propuesta de análisis factorial de datos *ipsatizados* basada en el modelo de Jackson y Alwin (1980) en el contexto del AFC. Los autores incluyeron en su trabajo un pequeño estudio de simulación para ejemplificar el procedimiento propuesto que demostraba que éste posee un funcionamiento adecuado, logrando recuperar el modelo de manera satisfactoria.

En un trabajo posterior, que incorpora el estudio de algunas condiciones, como el número de variables (6, 9 y 12), el tamaño de la muestra (100, 250 y 500), y el grado de *ipsatividad* de los datos (nula, total y parcial), Chan y Bentler (1996) aportaron hallazgos importantes acerca del funcionamiento del AFC de datos *ipsatizados*. En primer lugar, al comparar los valores de χ^2 en modelos con datos *normativos* e *ipsatizados*, observaron que el ajuste del modelo *ipsativo* era algo más pobre (Chan y Bentler, 1993; 1996). Sin embargo, al comparar las proporciones de rechazo de la hipótesis de ajuste con ambos tipos de datos, los índices de acuerdo mostraron una correspondencia que superaba el 80% en todas las condiciones. Esto implica que los modelos analizados con datos *ipsatizados* ajustan satisfactoriamente cuando el modelo se encuentra correctamente especificado. Sin perjuicio de lo anterior, estos estudios no analizaron otros índices de bondad de ajuste, ni la potencia del contraste del estadístico χ^2 bajo condiciones de falta de ajuste. En segundo lugar, hallaron que las estimaciones obtenidas con el modelo ajustado con datos *ipsatizados* alcanzaban valores muy cercanos a los obtenidos con los datos *normativos* antes de ser transformados (Chan, 2003; Chan y Bentler, 1993, 1996; Cheung, 2004). Estos estudios muestran que cuando el modelo está correctamente especificado, la recuperación de los parámetros con datos *ipsatizados* es satisfactoria.

A pesar de que los parámetros se recuperan adecuadamente, los errores de estimación se ven seriamente afectados. Esto se debe a la eliminación de una variable para evitar la singularidad de la matriz de covarianzas de los datos *ipsativos*, lo que lleva necesariamente a la pérdida de información para el proceso de estimación de los

parámetros, y con ello, a la obtención de estimadores eficientes pero con pérdida significativa de la precisión (Chan y Bentler, 1993, 1996; Cheung, 2004).

Por último, y como hemos mencionado anteriormente, aunque el grado de acuerdo en la decisión sobre la hipótesis de ajuste entre datos *normativos* e *ipsativos* sobrepasaba el 80% en todas las condiciones, este porcentaje aumentaba a medida que aumentaba también el número de ítems en el modelo (por ejemplo, el porcentaje de acuerdo sobrepasaba el 96% en todas las condiciones con 12 variables). Estos resultados indican que la evaluación del ajuste del modelo mejora significativamente a medida que aumenta el número de variables debido a que, al eliminar una variable de la matriz de entrada, la pérdida de información es proporcionalmente menor a medida que aumenta el número de variables en el modelo.

3.2. EL ERROR DE ESPECIFICACIÓN EN EL AFC DE VARIABLES IPSATIZADAS

Recientemente, Ximénez y Calderón (2012) han publicado un trabajo en el cual analizaron la sensibilidad de algunos índices de ajuste en la detección del error de especificación en el AFC con datos *ipsatizados*. Los autores, además de evaluar otros factores que ya habían sido estudiados (e.g., tamaño de la muestra, número de ítems y tipo de datos), estudiaron el error de especificación del modelo por infra-factorización, un error hallado comúnmente en la práctica investigadora.

Basándose en el estudio de Chan y Bentler (1996), Ximénez y Calderón (2012) analizaron el efecto de cuatro variables independientes: 1) el tipo de datos (*normativos* versus *ipsatizados*), 2) el tamaño de la muestra (100, 300, 500 y 1.000

observaciones), 3) el número de variables en el modelo (9, 12 y 18), y 4) el tipo de especificación en el modelo (*correcta* versus *incorrecta*). La última constituye una condición que no había sido examinada en la investigación previa. La forma de introducir error en el modelo fue mediante la omisión de un factor. Se eligió esta condición porque la investigación previa ha demostrado que es un tipo de error que produce una gran distorsión en la recuperación del modelo (Fava y Velicer, 1996; Ximénez, 2009). La figura 6 muestra el diseño planteado en el estudio.

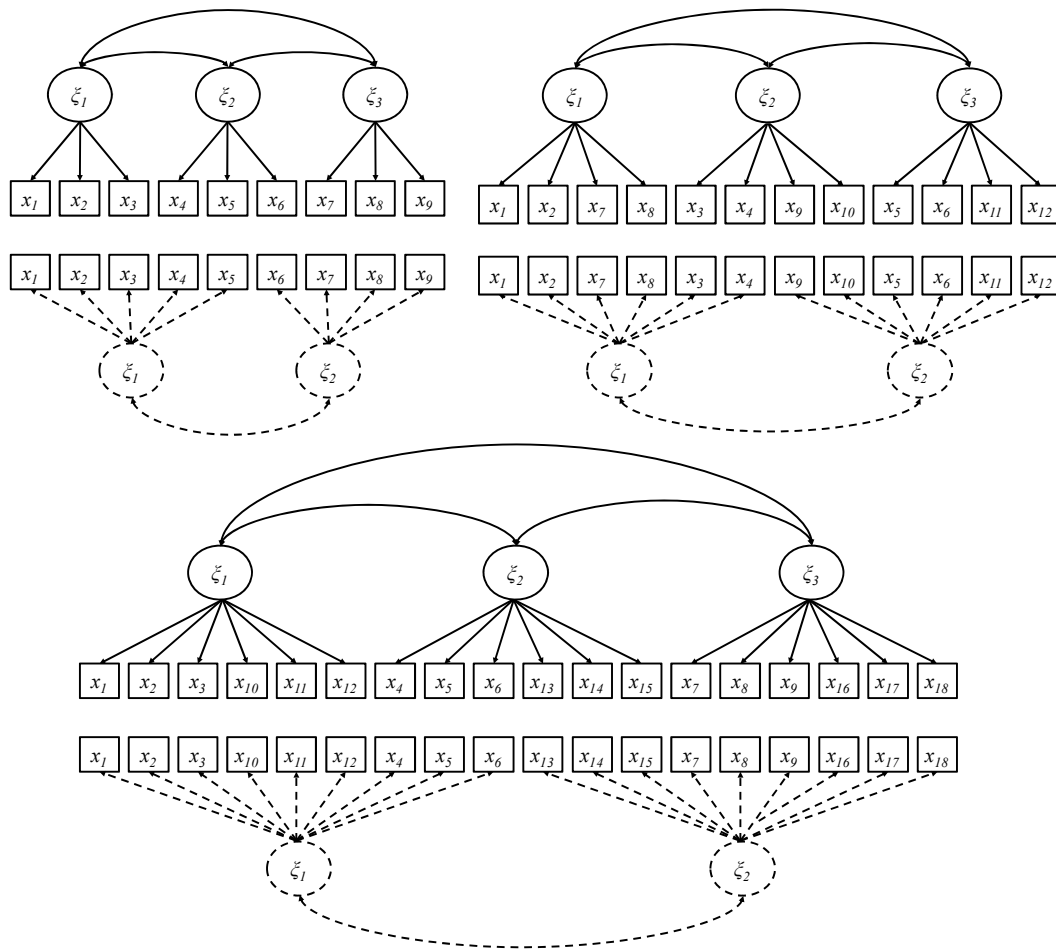


Figura 6. Diseño de los modelos estimados en el estudio de Ximénez y Calderón (2012).

El modelo correctamente especificado aparece representado en línea continua en el panel superior y el incorrecto en línea discontinua en el panel inferior (tomado de Ximénez y Calderón, 2012, pág. 306).

Al igual que en el estudio de Chan y Bentler (1996), los parámetros se estimaron mediante el método *ML* y la varianza de los factores fue fijada a 1 para la identificación del modelo. Las variables dependientes estudiadas correspondieron a la recuperación de las saturaciones evaluada mediante el índice de congruencia, *IC* de Tucker (1951), los errores estimación, la bondad de ajuste (*RMSEA*) y la potencia del contraste de la hipótesis de ajuste (proporción empírica de rechazo del test χ^2).

Ximénez y Calderón (2012) encontraron resultados similares a los obtenidos por Chan y Bentler (1996) en cuanto a los modelos correctamente especificados. Globalmente, en los modelos correctamente especificados, el funcionamiento del modelo *ipsativo* fue similar al del modelo *normativo*, tanto en los indicadores de ajuste como en la recuperación de los parámetros. No así en cuanto a los errores de estimación, los cuales, como cabía esperar, se vieron significativamente alterados debido a la información perdida al eliminar una variable de la matriz de covarianzas para evitar su singularidad.

En cambio, en modelos con error de especificación el funcionamiento del modelo *ipsativo* empeoró significativamente. Concretamente, frente al error de especificación, los índices de bondad de ajuste tuvieron dificultades para detectar el error con datos *ipsatizados*. La tabla 1 muestra los resultados del índice *RMSEA* obtenidos por los autores en cada una de las condiciones de estudio. Como se observa, cuando el modelo está correctamente especificado, los resultados con datos *ipsatizados* fueron muy similares a los datos *normativos* en todas las condiciones analizadas. Por el contrario, frente al error de especificación, el modelo *ipsatizado*

alcanzó valores significativamente más bajos que el modelo *normativo*. Esto nos muestra las dificultades que posee el estadístico *RMSEA* para detectar el error de especificación con datos *ipsatizados*.

Tabla 1.
Estadísticos descriptivos del índice de ajuste *RMSEA* obtenidos por Ximénez y Calderón (2012) en cada una de las condiciones del estudio (pág. 306)

N	P	Modelo correcto				Modelo incorrecto			
		Normativo		Ipsatizado		Normativo		Ipsatizado	
		M	SD	M	SD	M	SD	M	SD
100	9	.02	.02	.02	.03	.14	.02	.10	.03
	12	.02	.02	.02	.02	.13	.02	.11	.02
	18	.03	.02	.02	.02	.11	.01	.09	.01
300	9	.01	.02	.01	.02	.14	.01	.11	.01
	12	.01	.01	.01	.01	.13	.01	.11	.01
	18	.01	.01	.01	.01	.11	.01	.08	.01
500	9	.01	.01	.01	.01	.14	.01	.11	.01
	12	.01	.01	.01	.01	.13	.01	.11	.01
	18	.01	.01	.01	.01	.11	.00	.08	.01
1000	9	.01	.01	.01	.01	.14	.01	.11	.01
	12	.00	.01	.00	.01	.13	.01	.11	.01
	18	.00	.01	.00	.01	.11	.00	.08	.00

Nota. N = tamaño de la muestra; P = número de ítems; M = media; SD = desviación típica.

La tabla 2 muestra los resultados del ANOVA factorial que reporta los efectos principales y de doble interacción considerados en el estudio. La figura 7 muestra en mayor detalle los resultados de los efectos más importantes. Los resultados anteriormente comentados se pueden ver en mayor detalle en el gráfico A de la figura 7, que muestra el efecto de la doble interacción E^*D , el cual corresponde al efecto más importante reportado por los autores ($\eta^2 = .92$). Como se observa, cuando el

modelo está correctamente especificado, el índice *RMSEA* alcanza valores muy similares con ambos tipos de datos. Por el contrario, cuando el modelo está incorrectamente especificado, el índice *RMSEA* alcanza valores significativamente menores con datos *ipsatizados* que con datos *normativos*. Los gráficos B y C de la figura 7 muestran las interacciones *D*I* y *D*N*, respectivamente, ambas en condiciones de error de especificación. Estos gráficos ponen de manifiesto que la diferencias en el ajuste del índice *RMSEA* entre los datos *normativos* e *ipsatizados* va disminuyendo en la medida que aumenta el tamaño de la muestra y el número de variables en el modelo.

Tabla 2.
Resultados de ANOVA del índice de ajuste *RMSEA* obtenidos por Ximénez y Calderón (2012)
para los efectos principales y de doble interacción (pág. 307)

	<i>gl</i>	<i>F</i>	<i>p</i>	η^2
<i>D</i>	1	77390.519	<.001	.932
<i>E</i>	1	153507.278	<.001	.965
<i>N</i>	3	1202.800	<.001	.390
<i>I</i>	2	1030.225	<.001	.268
<i>D*E</i>	1	61299.765	<.001	.916
<i>D*N</i>	3	39.816	<.001	.021
<i>D*I</i>	2	29.289	<.001	.010
<i>E*N</i>	3	277.648	<.001	.129
<i>E*I</i>	2	513.776	<.001	.154
<i>N*I</i>	6	15.902	<.001	.017

Nota. *D*= tipo de dato; *E*=especificación; *N*=tamaño de la muestra; *I*=N° de variables.

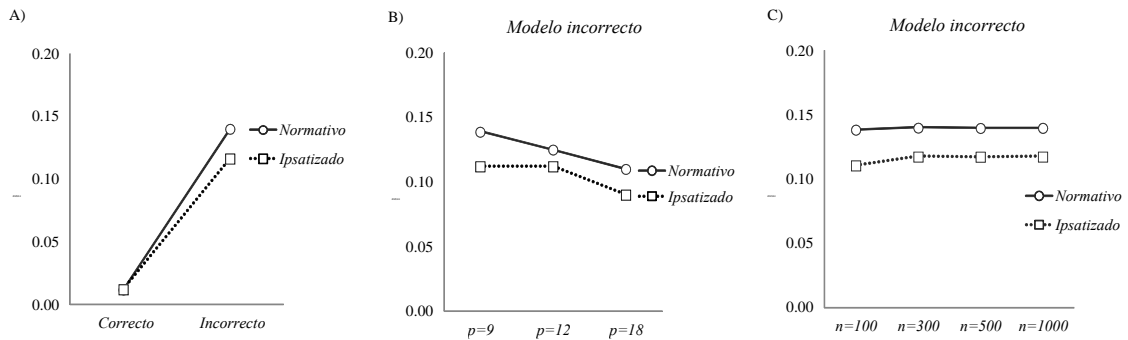


Figura 7. Gráficos que representan los efectos de interacción más importantes obtenidos (tomado de Ximénez y Calderón, 2012, pág. 307).

Los resultados del estudio de Ximénez y Calderón (2012) relativos a los problemas de los índices de ajuste para detectar el error de especificación con datos *ipsatizados*, sugieren recomendaciones prácticas para los investigadores que *ipsatizan* sus datos para evitar el sesgo de respuesta, que implican ser cautelosos en caso de tener sospecha de la existencia de errores de especificación por infra-factorización, y en especial, cuando el número de variables sea reducido y la muestra pequeña.

3.3. SENSIBILIDAD DE LOS ÍNDICES DE AJUSTE A LA FALTA INVARIANZA FACTORIAL

En cuanto al análisis de invarianza factorial, no existen hasta el momento estudios que evalúen el desempeño de esta técnica con datos *ipsatizados*. Sin embargo, creemos que la evidencia existente con datos *normativos* puede ser de utilidad para poder identificar algunos factores que afecten a la sensibilidad de los índices de ajuste en el análisis de invarianza factorial con datos *ipsatizados*.

Con datos *normativos* se han identificado varios factores que afectan a la prueba de la invarianza factorial. A nivel configural, Kaplan (1989) es el único que ha estudiado el funcionamiento de las pruebas de ajuste en condiciones de no invarianza configural. En su trabajo, el error de especificación fue introducido añadiendo ítems con saturaciones factoriales en dos factores en el grupo no invariante. Los resultados mostraron que la potencia fue más alta cuando las communalidades de los ítems en los factores fueron altas y cuando las saturaciones eran grandes, lo que sugiere que la potencia de las pruebas de bondad de ajuste estaría relacionada con la presencia de altas communalidades en el grupo no invariante.

En cuanto a la invarianza métrica, los hallazgos más importantes están relacionados con el efecto del número de ítems que muestran diferencias en las saturaciones factoriales, así como la dirección de las diferencias simuladas (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004). Los resultados de estos estudios muestran que la falta de invarianza se detecta más fácilmente cuando el patrón de diferencias es mixto (las diferencias simuladas eran mayores y otras menores en el grupo no invariante) y cuantos más ítems simulados presentaron diferencias (Meade y Lautenschlager, 2004; Chen, 2007). Por otro lado, las estimaciones de las diferencias en las saturaciones factoriales fueron más precisas toda vez que el tamaño de la muestra aumentaba, mayor fue la sobre-determinación (mayor número de ítems por factor), y mayores fueron las communalidades (Meade y Bauer, 2007). Por el contrario, la sensibilidad de los índices de ajuste se vio seriamente afectada cuando el patrón de no invarianza era homogéneo (cuando las diferencias simuladas eran todas menores en el grupo no invariante). Es así que el

patrón de no invarianza es el factor que ha mostrado un mayor efecto sobre la sensibilidad de las medidas de ajuste.

En cuanto a los índices de ajuste en particular, los estudios han demostrado que algunos de ellos poseen un rendimiento diferencial dependiendo de las condiciones del modelo y el nivel de invarianza en el que sean utilizados. Por ejemplo, el índice *CFI* posee el mejor rendimiento debido a que ha mostrado ser independiente de la complejidad del modelo, del tamaño de la muestra, y del nivel de invarianza que se analice (Cheung y Rensvold, 2002). Al comparar los índices en distintos niveles de invarianza, *CFI* y *RMSEA* han mostrado mayor sensibilidad para detectar las diferencias en las saturaciones factoriales y mayor consistencia a través de los distintos niveles de invarianza. Mientras que *SRMR* ha mostrado diferentes niveles de sensibilidad dependiendo del nivel de invarianza, siendo más sensible a las diferencias en la saturaciones factoriales, aunque se ve considerablemente afectado por el tamaño de la muestra (Chen, 2007).

En términos globales, el rendimiento general de los índices de ajuste está directamente relacionado con la sensibilidad que tengan a las variaciones aleatorias. En un extenso estudio de simulación, Chen (2007) demostró que existe una menor variación muestral de los índices de ajuste en muestras grandes. Esto sugiere que bajo la hipótesis nula, es más fácil cometer un error tipo I cuando la muestra es pequeña. Mientras que bajo la hipótesis alternativa, es más fácil cometer un error tipo II cuando la muestra es grande. Como señalan Hu y Bentler (1999), un buen criterio para evaluar el ajuste de un modelo debe minimizar el error Tipo I y el error Tipo II, simultáneamente.

En un esfuerzo en esta dirección, Chen (2007) propone una serie de criterios de corte para varios índices de ajuste y para distintos tipos de invarianza. Particularmente, para el nivel de invarianza métrico sugiere valores de cambio del ajuste de entre $-.005$ ó $-.010$ para *CFI*; entre $.015$ y $.010$ para *RMSEA*; y entre $.025$ y $.030$ para *SRMR*.

A pesar de la potencial ventaja que posee la *ipsatización* para evitar el impacto del sesgo de respuesta en estudios de invarianza de medida, es notable la falta de estudios que intenten evaluar el funcionamiento del AFCM con variables *ipsatizadas*. Conocer el funcionamiento general y determinar las condiciones óptimas en las cuales es posible realizar un AFCM de variables *ipsatizada*, permitirá poder avanzar en el estudio de la capacidad de esta estrategia para tratar el sesgo de respuesta.

Aunque existen alternativas al tratamiento del sesgo de respuesta, como lo es el enfoque instrumental, alguno de estos procedimientos puede tener altos costes desde un punto de vista práctico y/o técnico. Por ejemplo, la implementación de sub-escalas para la detección de patrones de respuesta *infra-dimensionados* o *sobre-dimensionados*, como es el caso del *Inventario Multifásico de Personalidad de Minnesota*, MMPI (Butcher, Dahlstrom, Graham, Tellegen y Kaemmer, 1989), aumenta considerablemente el tiempo de administración y contribuye al agotamiento de los entrevistados, afectando a la fiabilidad de las medidas. Además, esta estrategia sólo permite la detección y posterior eliminación de los casos, y en ningún caso el tratamiento del sesgo. Por otro lado, la utilización de formatos de respuesta para evitar el *sesgo de respuesta*, como es el caso de los test de *respuesta forzada*, puede

generar problemas de comprensión por parte de los entrevistados, especialmente cuando los bloques contienen ítems opuestos. Adicionalmente, como sugieren Brown y Maydeu-Olivares (2011), este tipo de test puede traer consigo problemas técnicos, durante el proceso de estimación o en la evaluación del ajuste, especialmente cuando los modelos son considerablemente grandes y/o poseen muchos grados de libertad.

Por tanto, el desarrollo de alternativas al tratamiento instrumental, y más concretamente, la *ipsatización* de variables, se convierte en una excelente alternativa al tratamiento del *sesgo de respuesta*. Esto se debe fundamentalmente a que la *ipsatización* permitiría la utilización de escalas *normativas*, específicamente la escala *Likert*, facilitando el proceso de respuesta de los sujetos y evitando la implementación de complejos dispositivos instrumentales. En consecuencia, el propósito del presente trabajo es evaluar el funcionamiento del AFCM con variables *ipsatizadas* en el contexto del análisis de E/IM, así como las condiciones óptimas en las cuales se debe llevar a cabo un análisis de invarianza factorial con datos *ipsatizados*.

CAPITULO IV

OBJETIVOS DE LA TESIS

En este capítulo se describen los objetivos de la tesis, el planteamiento de las hipótesis, y la justificación para el planteamiento de los tres estudios que hemos llevado a cabo para el análisis de los objetivos propuestos.

4.1. OBJETIVOS

Dada la capacidad potencial de la *ipsatización* de variables *normativas* para evitar el impacto del *sesgo de respuesta*, y que el AFCM es una de las herramientas más utilizadas en el estudio de la E/IM entre distintos grupos, se hace necesario conocer el funcionamiento del análisis de invarianza factorial con variables *ipsatizadas*. Por tanto, el ***objetivo general*** del presente trabajo es examinar el funcionamiento de la *ipsatización* en el tratamiento del *sesgo de respuesta* y evaluar la sensibilidad de los índices de bondad de ajuste más comúnmente utilizados en la evaluación del ajuste de modelos de invarianza factorial con variables *ipsatizadas*. Conocer el funcionamiento del AFCM con variables *ipsatizadas* en general y de los índices de bondad de ajuste en particular, permitiría proponer el uso de la *ipsatización* como estrategia para evitar el impacto del *sesgo de respuesta* y realizar recomendaciones prácticas acerca de cómo llevar a cabo un análisis de E/IM con variables *ipsatizadas*.

En cuanto a los ***objetivos específicos***, son tres:

- 1) En primer lugar, nuestro análisis se centra en la capacidad para detectar la falta de invarianza debido a que la literatura previa revela los problemas que poseen los índices de ajuste para detectar el error de especificación con variables *ipsatizadas* (Ximénez y Calderón, 2012). Como ha sido mencionado anteriormente, mientras que el funcionamiento general de AFC con variables *ipsatizadas* es similar al de variables *normativas* en modelos correctamente especificados, los índices de ajuste pierden capacidad para detectar la falta de ajuste en modelos incorrectamente especificados, tornándose excesivamente conservadores.
- 2) En segundo lugar, el trabajo se focaliza en la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza factorial a nivel configural y métrico. A pesar de la existencia de diferentes niveles de invarianza, nos hemos centrado en estos dos niveles, ya que el cumplimiento de la invarianza a nivel configural y métrico es condición necesaria y suficiente para poder determinar la E/IM cuando el interés se centra en la validez de constructo, especialmente cuando el propósito es la comparación entre distintos grupos (Dimitrov, 2010).
 - A nivel configural, el propósito es evaluar la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza por infra-factorización. La literatura previa muestra que este tipo de error genera importantes distorsiones en la recuperación del modelo factorial con variables *normativas* (Fava y Velicer, 1996; Ximénez, 2009), y que las medidas de bondad de ajuste en AFC

de variables *ipsatizadas* se ven seriamente afectadas (Ximénez y Calderón, 2010).

- En cuanto al nivel métrico, el propósito es evaluar la sensibilidad de los índices de bondad de ajuste para detectar las diferencias en las saturaciones factoriales. En este nivel hemos considerado la dirección de las diferencias, ya que los estudios previos con variables *normativas* muestran que el patrón de la dirección de las diferencias entre las saturaciones factoriales de los diferentes grupos tienen un efecto significativo sobre la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza a nivel métrico (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004).
- 3) Por último, debido a la ausencia de investigación previa que evalúe la efectividad de la *ipsatización* en el tratamiento de *sesgo de respuesta*, otro de los objetivos del presente trabajo es comparar el AFCM de variables *ipsatizadas* con otra de las estrategias propuestas para evitar el impacto de este tipo respuesta en el estudio de la E/IM entre distintos grupos: los test de formato de *respuesta forzada*, y su análisis a través del modelo AFC Thurstoniano de Brown y Maydeu-Olivares (2011). Este formato de respuesta constituye una de las estrategias más utilizadas desde el punto de vista instrumental en el tratamiento del *sesgo de respuesta*, ya que contribuye a eliminar la *aquiescencia* (Cheung y Chan) y a reducir el efecto halo (Bartram, 2007), que son algunas de las fuentes más importantes de *sesgo de respuesta*. La comparación del funcionamiento de ambas estrategias puede contribuir a determinar si la *ipsatización* de variables logra eliminar el impacto del *sesgo de respuesta* en el análisis de la E/IM.

4.2. PLAN DE TRABAJO

En cuanto al plan de trabajo para abordar los objetivos propuestos, hemos planteado dos tipos de estudios:

- 1) En primer lugar, hemos llevado a cabo dos *estudios de simulación*, los cuales tienen como propósito evaluar el funcionamiento de los índices de ajuste de los modelos AFCM con datos *ipsatizados*, así como determinar las condiciones óptimas para llevar a cabo un análisis de invarianza factorial con este tipo de datos. Ambos estudios están organizados sistemáticamente en función del procedimiento convencional del análisis de invarianza factorial.
 - El primer estudio se centra en el análisis del ajuste del modelo configural o base. En este primer estudio, la falta de invarianza ha sido introducida a través del error de especificación por infra-factorización en uno de los grupos (omitiendo un factor en el grupo no invariante), debido a que este tipo de error de especificación genera importantes distorsiones en la recuperación del modelo (Fava y Velicer, 1996; Ximénez, 2006, 2009) y porque ha sido el único error a nivel configural que ha sido estudiado con datos *ipsatizados* (Ximénez y Calderón, 2012).
 - El segundo estudio tiene como objetivo evaluar el funcionamiento de los índices de ajuste en la detección de la falta de invarianza en las saturaciones factoriales de los distintos grupos, asumiendo la existencia de invarianza configural. Hemos introducido la falta de invarianza a través de dos condiciones: el patrón de *no invarianza homogéneo* y el patrón de *no invarianza mixto*, ya que, como se ha hallado en la investigación previa, la

dirección o patrón de las diferencias en las saturaciones factoriales ha mostrado ser uno de los factores con mayor efecto sobre los resultados de los índices de bondad de ajuste en este nivel. A diferencia del estudio 1, en el cual evaluamos los valores absolutos de los índices de ajuste, en el estudio 2, puesto que el objetivo es analizar la pérdida en el ajuste, hemos considerado la diferencia entre los índices de bondad de ajuste correspondientes al nivel configural y al nivel métrico.

Ya que el propósito fundamental de la *ipsatización* de variables es evitar el impacto del *sesgo de respuesta* en estudios de E/IM, no hemos considerado niveles de invarianza más restrictivos (interceptos, varianza de los errores, varianzas de los factores, etc.). Adicionalmente, el nivel de invarianza métrico es requisito suficiente para considerar a una medida equivalente, cuando el interés está centrado en la validez de constructo y en la comparación entre distintos grupos. En cuanto a los demás niveles de medida, por un lado, la invarianza escalar carece de sentido debido a que el proceso de *ipsatización* comprende una estandarización de las variables con respecto a la media de cada sujeto, por tanto el valor de los interceptos no tiene interpretación. Y por otro, la invarianza de los errores de medida es un requisito demasiado exigente, muy difícil de cumplir en la práctica investigadora y muy pocas veces considerado por los investigadores, toda vez que carece de interés en la comparación entre las medias del constructo entre diferentes grupos (Bentler, 2004; Byrne, 1988).

En ambos estudios hemos considerado otro tipo de factores que han sido evaluados tanto en el funcionamiento del análisis de invarianza factorial con

variables *normativas*, así como estudios que evalúan el funcionamiento del AFC con variables *ipsatizadas*, como son: el número de ítems y el tamaño de la muestra. Como variables dependientes hemos considerado los índices de ajuste más utilizados en la práctica investigadora, y que han sido estudiados tanto en el AFC de variables *ipsatizadas*, como en el AFCM de variables *normativas*; el test de bondad de ajuste a través del estadístico χ^2 , los índices de ajuste absolutos *RMSEA* (Steiger y Lind, 1980) y *SRMR* (Bentler, 1990), y el índice de ajuste incremental *CFI* de Bentler (1990).

- 2) En segundo lugar, hemos realizado un *estudio empírico*, que tiene dos objetivos. El primero es ilustrar el procedimiento del AFCM con variables *ipsatizadas* a través del método *DE* de Cheung y Chan (2004) y su implementación en un programa SEM. El segundo objetivo es evaluar el funcionamiento del AFCM de variables *ipsatizadas* con datos empíricos. Hasta la fecha, no existen estudios concluyentes con respecto a la efectividad de la *ipsatización* de variables en el tratamiento del *sesgo de respuesta* en estudios de E/IM. Por tanto, nuestro estudio tiene carácter exploratorio. Hemos estudiado el funcionamiento y efectividad de la *ipsatización* en la eliminación del impacto del sesgo de respuesta en estudios de AFCM, a través de método *DE* de Cheung (2004), considerando las condiciones óptimas que hemos determinado tras los dos estudios de simulación previos. Para ello, hemos obtenido los datos utilizando una versión reducida del “Inventario de Estilos de Aprendizaje” de Kolb (1985), un instrumento ampliamente utilizado en el ámbito educativo. El test fue

presentado en dos formatos: el formato original que consiste en un test de *respuesta forzada*, y una versión *normativa* presentada en formato Likert. Para evaluar el desempeño de la *ipsatización* hemos comparado los resultados obtenidos del ajuste de tres conjuntos de datos: 1) un conjunto de datos *normativos* obtenido a partir del formato Likert, el cual hemos analizado a través de un AFCM convencional; 2) el AFCM utilizando el método *DE* sobre la transformación *ipsativa* de los datos *normativos*; y 3) hemos analizado los datos del formato de *respuesta forzada*, utilizando el modelo Thurstoniano de AFC de Brown y Maydeu-Olivares (2011).

4.3. HIPÓTESIS

Aunque en la literatura previa no existen trabajos que traten específicamente el estudio del AFCM con variables *ipsatizadas* y su funcionamiento en el tratamiento en el *sesgo de respuesta* en el contexto del análisis de E/IM, los hallazgos existentes de estudios previos con respecto al AFC de variables *ipsatizadas* y al AFCM con variables *normativas*, nos permiten plantear algunas hipótesis con respecto a los resultados esperados de nuestros estudios. A continuación se listan cada una de las hipótesis:

- 1) Debido a que la investigación previa con respecto al AFC de datos *ipsatizados* ha mostrado que los **índices de bondad de ajuste** poseen un comportamiento similar a los índices del AFC de datos *normativos* (Chan, 2003; Cheung, 2004; Chan y Bentler, 1993, 1996, Ximénez y Calderón, 2012), no esperamos encontrar diferencias estadísticamente significativas en los índices de bondad de ajuste

entre datos *normativos* e *ipsatizados* en el AFCM cuando los grupos son invariantes. Por el contrario, dado que los índices de ajuste sufren una pérdida significativa de la sensibilidad para detectar el error de especificación en el AFC (Ximénez y Calderón, 2012), cabe esperar que en el AFCM los índices de bondad de ajuste también sufran una pérdida significativa de sensibilidad para detectar la falta de invarianza entre los grupos. Este fenómeno ocurrirá tanto en el nivel de invarianza configural como en el nivel de invarianza métrico.

- 2) En cuanto a la **falta de invarianza a nivel métrico**, la literatura previa, con respecto al AFCM de variables *normativas*, muestra que el patrón de diferencias en las saturaciones factoriales posee un efecto estadísticamente significativo en la sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza. Los resultados de estos estudios muestran que la pérdida de sensibilidad de los índices de bondad de ajuste es mayor cuando el patrón de las diferencias es homogéneo que cuando es mixto (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004). Por tanto, esperamos que, pese a que la pérdida de sensibilidad en el AFCM de variables *ipsatizadas* ocurrirá frente a ambos patrones de no invarianza, ésta será mayor cuando el patrón de las diferencias de las saturaciones factoriales entre los grupos sea homogénea que cuando sea mixta.
- 3) En cuanto al **número variables en el modelo**, a diferencia de lo que ocurre con variables *normativas*, los estudios anteriores muestran que el número de ítems en el modelo parece ser un factor importante en el AFC de variables *ipsatizadas*.

Esto se debe a que en el proceso de factorización de variables *ipsatizadas* existe una pérdida de información debido a la necesaria la eliminación de una fila y una columna de la matriz de covarianzas de entrada para evitar su singularidad (Cheung, 2004; Chan y Bentler, 1996). Junto con ello, los hallazgos encontrados por Ximénez y Calderón (2012) demostraron que el número de variables en el modelo posee un efecto estadísticamente significativo sobre la sensibilidad de los índices de bondad de ajuste para detectar el error de especificación. Concretamente, estos autores demostraron que la pérdida de sensibilidad de los índices de bondad ajuste es mayor cuanto menor es el número de ítems en el modelo. Es por ello que esperamos que el número de variables tenga un efecto en la sensibilidad de los índices de ajuste. Más concretamente, en la medida que aumente el número de variables en el modelo, la pérdida de información será proporcionalmente menor. Por tanto, esperamos que la pérdida de sensibilidad de los índices de ajuste en el AFCM de variables *ipsatizadas* sea mayor en la medida en que el número de variables presentes en el modelo sea menor.

- 4) En cuanto a los resultados del *estudio empírico*, a grandes rasgos existen dos posibles escenarios:
 1. Por un lado, tal como sugieren Cheung y Chan (2002), al comparar los resultados de la bondad de ajuste entre los modelos con datos *normativos* e *ipsatizados*, podremos encontrar que las medidas de bondad de ajuste serán muy similares entre ambos tipos de datos. Estos resultados estarían indicando que los datos no se encuentran contaminados con la presencia de

sesgo de respuesta. Ello independientemente de que los índices de bondad de ajuste muestren la existencia o no de invarianza entre los grupos.

2. En segundo lugar, es posible que los resultados en el ajuste de ambos modelos no sean concordantes. Cabe esperar que, mientras que con datos *normativos*, los índices de bondad de ajuste indiquen la falta de invarianza entre los grupos, el modelo con datos *ipsatizados* indicará una mejora en el ajuste. Este resultado puede tener dos interpretaciones distintas: 1) o bien la diferencia en el ajuste entre ambos tipos de datos se debe a que existe un impacto del sesgo de respuesta y que la *ipsatización* logra eliminar su efecto, lo que nos lleva a concluir que los grupos son invariantes (Cheun y Chan, 2002); 2) o bien que los grupos son no invariantes y los índices de bondad de ajuste no poseen la sensibilidad adecuada para detectar la falta de invarianza. En este último escenario utilizaremos los resultados obtenidos con los datos de *respuesta forzada*. Un ajuste adecuado del modelo con este tipo de datos será indicador de la efectividad de la *ipsatización* para eliminar el sesgo de respuesta en el análisis E/IM, mientras que una falta de ajuste del modelo de *respuesta forzada*, sería indicador de que la *ipsatización* no logra evitar el *sesgo de respuesta* y que los resultados obtenidos se deben a la pérdida de sensibilidad de los índices de bondad de ajuste para detectar la falta de invarianza con datos *ipsatizados*.

CAPÍTULO V

ESTUDIOS DE SIMULACIÓN

En este capítulo se presenta el planteamiento y los resultados de los dos estudios de simulación llevados a cabo para el análisis de los objetivos propuestos. Adicionalmente, se expone una síntesis de los resultados obtenidos y se discuten las implicaciones prácticas de los mismos.

5.1. ESTUDIO 1: SENSIBILIDAD DE LOS ÍNDICES DE AJUSTE A LA INVARIANZA CONFIGURAL EN AFCM CON DATOS IPSATIZADOS

El propósito del primer estudio es evaluar la sensibilidad de algunos índices de bondad de ajuste para detectar la falta de *invarianza configural* en el AFCM de variables *ipsatizadas*. Del mismo modo que en Ximénez y Calderón (2012), hemos considerado la no invarianza en la estructura factorial introduciendo errores de especificación mediante la alteración del número de factores.

5.1.1. Diseño

El diseño del primer estudio sólo incluye el análisis de la *invarianza configural* entre dos grupos por ser el modelo más sencillo y porque los resultados obtenidos pueden ser fácilmente generalizables a condiciones de más de dos grupos.

Para simular la condición de *no invarianza configural* hemos diseñado una estructura factorial distinta en uno de los grupos que se comparan. La diferencia en la estructura factorial ha sido introducida añadiendo un factor al grupo no invariante (ver figura 8). Más concretamente, el grupo 1 (*grupo invariante*) ha sido simulado con una estructura de dos factores y el grupo 2 (*grupo no invariante*) con una estructura de tres factores. Posteriormente, al analizar la invarianza factorial, hemos especificado en la estimación de parámetros sólo dos factores en ambos grupos, cumpliendo así con el error de especificación por infra-factorización definido en el grupo 2.

El diseño contempla el estudio de cinco *variables independientes* (ver tabla 3). Más concretamente:

1) Tipo de invarianza configural (I): se define en dos niveles:

1. *Invarianza*: en esta condición hemos simulado una estructura de dos factores idéntica en ambos grupos (ver *modelo 1* en la parte superior de la figura 8).
2. *No invarianza* entre los grupos: esta condición ha sido introducida a través del error de especificación por infra-factorización. La estructura del grupo 2 ha sido modificada incluyendo un tercer factor (ver *modelo 2* e la parte inferior de la figura 8), manteniendo la estructura de dos factores para el grupo 1.

Para ambas condiciones, hemos especificado en el análisis un modelo de dos factores con el propósito de establecer el error de especificación por infra-factorización en el grupo 2 de la condición de *no invarianza*.

2) Tipo de datos (D): con el propósito de comparar los resultados de ajustar un modelo de invarianza con datos *ipsatizados* hemos implementado los análisis con

datos *normativos* (*pre-ipsativos*). Por tanto, este factor quedó constituido por dos niveles. Primero se generaron los datos *normativos*, que posteriormente fueron *ipsatizados* a través de la ecuación [1.10].

- 3) Nº de variables (P). Del mismo modo que en estudios anteriores, hemos evaluado el efecto del número de variables en el modelo (9, 12 y 18 ítems). Como hemos visto, el número de ítems ha mostrado tener un efecto significativo en los índices de bondad de ajuste en modelos con datos *ipsatizados*.
- 4) Tamaño de la muestra (N). Al igual que en los estudios de Chan y Bentler (1996) y Ximénez y Calderón (2012) hemos utilizado cuatro tamaños muestrales: 100, 300, 500 y 1000. Los primeros tres niveles reflejan los tamaños más utilizados en la práctica investigadora. El último nivel (1000), corresponde a una condición óptima, y servirá como criterio de comparación.
- 5) Nivel de invarianza (J). Ya que el interés de nuestro estudio se enmarca en el contexto del análisis de invarianza como evidencia de la validez de constructo en estudios comparativos, para cada una de las condiciones del estudio hemos ajustado los modelos tanto en el nivel *configural* como en el nivel *métrico*, para así estudiar las diferencias en los indicadores de ajuste.

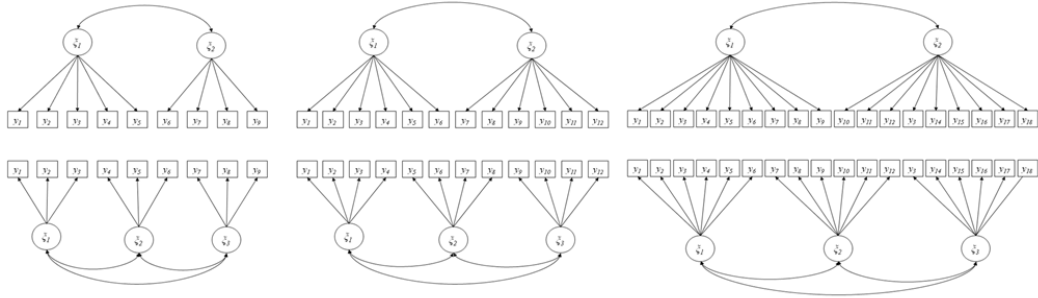
En cuanto a las **variables dependientes**, para estudiar la sensibilidad del modelo *ipsatizado* en la detección de falta de *invarianza configural*, hemos considerado la proporción empírica de rechazo de la hipótesis de ajuste (*PER*) según el estadístico χ^2 , además de los índices de ajuste que se utilizan con mayor frecuencia en la

investigación aplicada: el Índice de Ajuste Comparativo *CFI* de Bentler (1990), el Error de Aproximación Cuadrático Medio o *RMSEA* (Steinger y Lind, 1980) y el Residuo Cuadrático Medio Estandarizado o *SRMR* (Bentler, 1990). Estos índices han sido seleccionados, además de por ser los más utilizados en el ámbito aplicado, por ser los más estudiados tanto en el análisis de invarianza factorial multi-grupo con datos *normativos*, como en el AFC de datos *ipsatizados* (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004; Ximénez y Calderón, 2012).

Tabla 3.
Variables consideradas en el estudio I

Código	Variable	Niveles
<i>Variables independientes</i>		
<i>I</i>	Tipo de invarianza configural	Invariante No invariante
<i>D</i>	Tipo de datos	Normativos Ipsatizados
<i>P</i>	N° de variables	9 12 18
<i>N</i>	Tamaño de la muestra	100 300 500 1000
<i>J</i>	Nivel de invarianza	Configural Métrico
<i>Variables dependientes</i>		
<i>PER</i>	Proporción empírica de rechazo de la hipótesis de ajuste	
<i>CFI</i>	Comparative Fit Index	
<i>RMSEA</i>	Root-Mean Square Error of Approximation	
<i>SRMR</i>	Standardized Root Mean Square Residual	

Modelo 1



Modelo 2

Figura 8. Estructura de los modelos teóricos simulados. Para simular la invarianza entre los grupos hemos establecido la estructura del modelo 1 en ambos grupos. Para la condición de “no invarianza” hemos establecido la estructura del modelo 1 para el grupo 1, y la del modelo 2 para el grupo 2.

5.1.2. Procedimiento

En primer lugar hemos definido las estructuras factoriales para el cálculo de las matrices de covarianzas poblacionales. Para definir las estructuras factoriales y los valores de los parámetros poblacionales nos hemos basado en los trabajos de Chan y Bentler (1996) y Ximénez y Calderón (2012). Los valores teóricos de los parámetros no estandarizados (ver anexo 4) corresponden a valores estandarizados de .707 para todas las saturaciones factoriales y a .50 para todos los errores de medida ($.707^2 + .50 = 1$). Para todas las condiciones, las correlaciones entre los factores han sido establecidas en .50, tal y como se ha hecho en estudios anteriores (Chan, 2003; Chan y Bentler, 1993; 1996; Ximénez y Calderón, 2012).

A partir de las matrices calculadas en base a los valores poblacionales, hemos generado 500 réplicas por condición, simuladas con distribución normal multivariante con media $\mu = 0$ y matriz de covarianzas $\Sigma = \Lambda\Phi\Lambda + \Psi$. Estos datos *normativos*

fueron transformados a partir de la ecuación [1.10], para obtener los correspondientes datos *ipsatizados*.

Una vez generados los datos, hemos realizado un AFC en cada una de las réplicas. Para el caso de los datos *normativos* hemos llevado a cabo un AFC convencional. Para los datos *ipsatizados* hemos utilizado el método *DE* de Cheung (ver anexo 2 para consultar la sintaxis utilizada). Con ambos tipos de datos hemos utilizado el método de estimación de *Máxima Verosimilitud* (ML) implementado en el programa EQS 6 (Bentler, 2006). Siguiendo la recomendación de Cheung (2004), para establecer la métrica de los factores latentes su varianza se fijó a 1.

5.1.3. Análisis de datos

En primer lugar, hemos realizado un análisis descriptivo de las variables dependientes para cada una de las condiciones del estudio. Para examinar el efecto de las variables independientes en cada una de las variables dependientes, hemos llevado a cabo un ANOVA factorial, y hemos utilizado el estadístico η^2 para determinar la significación práctica de los efectos que resulten significativos. El modelo factorial contempla dos variables intra-sujeto; (*D*) el *Tipo de datos* (*normativo* vs. *ipsatizado*) y (*J*) el *Nivel de invarianza* (*configural* vs. *métrico*); y tres variables inter-sujeto: (*I*) el *Tipo de invarianza* (*invariante* vs. *no invariante*); (*P*) el *Nº de variables en el modelo* (9, 12 y 18), y (*N*) el *Tamaño de la muestra* (100, 300, 500 y 1000). El diseño del estudio quedó compuesto por $(D=2) \times (I=2) \times (J=2) \times (P=3) \times (N=4) = 96$ condiciones. El meta-modelo incorpora tanto los efectos principales, como los efectos de doble y triple interacción. A pesar de que Skrondal (2000) sugiere sólo el análisis

de las dobles interacciones, debido a las dificultades de interpretar interacciones de orden superior, hemos decidido analizar las triples interacciones desde un punto de vista gráfico, tal como sugieren León y Montero (2001, 2003), debido al interés del posible efecto de la interacción $D * I * P$ que hemos hipotetizado y a los efectos encontrados en este nivel de interacción en la literatura previa.

5.1.4. Resultados

Las tablas 4 y 5 muestran la proporción de soluciones inadecuadas en cada una de las condiciones de estudio. Del total de soluciones (48.000), sólo un 3.5% no lograron la convergencia, las cuales corresponden principalmente a las soluciones ajustadas con datos *ipsatizados*. Las soluciones no convergentes disminuyeron a medida que el tamaño de la muestra y el número de variables en el modelo aumentaba.

En cuanto a los *Casos Heywood*, se observa un patrón claro. Mientras que el modelo *normativo* prácticamente no genera casos *Heywood*, los datos *ipsatizados* generan un importante número de soluciones con valores no admisibles, llegando incluso al 50% en algunas condiciones (e.g., $P=9$ y $N=100$). Este porcentaje va disminuyendo considerablemente, hasta llegar a 0, a medida que aumentan, tanto el número de variables en el modelo ($P=18$), como el tamaño de la muestra ($N \geq 300$). Los Casos Heywood disminuyeron considerablemente en las condiciones de restricciones de igualdad en las saturaciones factoriales.

Tabla 4.
Proporción soluciones no convergentes en cada una de las condiciones del estudio

P	N	Invariante				No invariante			
		Normativo		Ipsatizado		Normativo		Ipsatizado	
		CF	MT	CF	MT	CF	MT	CF	MT
P=9	100	.000	.000	.018	.034	.000	.000	.012	.010
	300	.000	.000	.006	.024	.000	.000	.004	.004
	500	.000	.000	.000	.024	.000	.000	.000	.002
	1000	.000	.000	.000	.032	.000	.000	.002	.000
P=12	100	.000	.000	.194	.086	.000	.000	.100	.040
	300	.000	.000	.084	.064	.000	.000	.038	.010
	500	.000	.000	.034	.020	.000	.000	.026	.002
	1000	.000	.000	.018	.012	.000	.000	.012	.000
P=18	100	.000	.000	.002	.000	.000	.000	.002	.000
	300	.000	.000	.000	.000	.000	.000	.000	.000
	500	.000	.000	.000	.002	.000	.000	.000	.000
	1000	.000	.000	.000	.002	.000	.000	.000	.000

Nota. CF = Nivel de invarianza configural; MT = Nivel de invarianza métrica

Tabla 5.
Proporción de Casos Heywood en cada una de las condiciones del estudio

P	N	Invariante				No invariante			
		Normativo		Ipsatizado		Normativo		Ipsatizado	
		CF	MT	CF	MT	CF	MT	CF	MT
P=9	100	.000	.000	.468	.122	.004	.004	.500	.234
	300	.000	.000	.166	.062	.000	.000	.094	.034
	500	.000	.000	.098	.064	.000	.000	.078	.014
	1000	.000	.000	.040	.082	.000	.000	.036	.000
P=12	100	.000	.000	.458	.146	.002	.000	.344	.084
	300	.000	.000	.312	.092	.000	.000	.186	.018
	500	.000	.000	.200	.036	.000	.000	.112	.006
	1000	.000	.000	.082	.014	.000	.000	.032	.000
P=18	100	.000	.000	.002	.000	.000	.000	.010	.000
	300	.000	.000	.000	.000	.000	.000	.000	.000
	500	.000	.000	.000	.002	.000	.000	.000	.000
	1000	.000	.000	.000	.002	.000	.000	.000	.000

Nota. CF = Nivel de invarianza configural; MT = Nivel de invarianza métrica

La proporción de soluciones no convergentes y el patrón de Casos Heywood son similares a los obtenidos por Chan y Bentler (1998) y Ximénez y Calderón (2012), aunque estos últimos sólo obtuvieron Casos Heywood para el modelo correctamente especificado en sólo una de las condiciones ($N=100$ y $P=9$). Las soluciones inadecuadas fueron eliminadas para los análisis posteriores.

La tabla 6 muestra la proporción empírica de rechazo (PER) de la hipótesis de ajuste según el estadístico χ^2 .

Tabla 6.
Proporción rechazo de la hipótesis de ajuste en cada una de las condiciones del estudio

<i>P</i>	<i>N</i>	<i>Invariante</i>				<i>No invariante</i>			
		<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>	
		<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>
<i>P=9</i>	<i>100</i>	.040	.056	.016	.032	.975	.990	.637	1.000
	<i>300</i>	.055	.048	.078	.060	1.000	1.000	1.000	1.000
	<i>500</i>	.040	.058	.054	.068	1.000	1.000	1.000	1.000
	<i>1000</i>	.063	.050	.059	.061	1.000	1.000	1.000	1.000
<i>P=12</i>	<i>100</i>	.133	.122	.106	.099	1.000	1.000	.946	.997
	<i>300</i>	.063	.057	.057	.042	1.000	1.000	1.000	1.000
	<i>500</i>	.041	.046	.056	.043	1.000	1.000	1.000	1.000
	<i>1000</i>	.042	.037	.035	.033	1.000	1.000	1.000	1.000
<i>P=18</i>	<i>100</i>	.253	.267	.222	.220	1.000	1.000	.978	1.000
	<i>300</i>	.106	.104	.096	.092	1.000	1.000	1.000	1.000
	<i>500</i>	.084	.084	.084	.084	1.000	1.000	1.000	1.000
	<i>1000</i>	.064	.066	.058	.068	1.000	1.000	1.000	1.000

Nota. CF = Nivel de invarianza configural; MT = Nivel de invarianza métrica

Cuando los grupos son invariantes, la proporción de rechazo se mantiene en torno al valor esperado (.05), con la excepción de las condiciones de $P=12$ y $N=100$, y $P=18$ y $N=100$. Ambos tipos de datos no presentan grandes diferencias en la proporción de rechazo del ajuste, tanto a nivel *configural* como a nivel *métrico*. Al

igual que los resultados reportados por Chan y Bentler (1996), el error tipo I aumenta a medida que aumenta la complejidad del modelo y disminuye el tamaño de la muestra. Para el caso de las condiciones de *no invarianza* entre los grupos, todas las condiciones se mantienen dentro de valores aceptables. Esto indica que el estadístico χ^2 no muestra una pérdida de la potencia en el contraste de la hipótesis de ajuste con datos *ipsatizados*.

La tabla 7 muestra los estadísticos descriptivos obtenidos para el índice *CFI* en cada una de las condiciones de estudio. Estos resultados concuerdan con los hallazgos anteriores en relación a las dificultades de los índices de ajuste para detectar el error de especificación con datos *ipsatizados*. Mientras que en condiciones de *invarianza* el comportamiento de *CFI* con datos *ipsatizados* es similar que con datos *normativos*, en condiciones de *no invarianza* éste cambia considerablemente entre ambos tipos de datos.

Es posible observar que en la condición de *no invarianza a nivel configural*, el índice *CFI* obtiene en promedio valores más altos con datos *ipsatizados* que con datos *normativos*. Estos incluso llegan a valores en torno a .95 en algunas condiciones (e.g., $P=9$), lo que indicaría un ajuste aceptable de la hipótesis de *invarianza configural*. Por otro lado, una vez ajustado el modelo de invarianza métrica, los valores de *CFI* en datos *ipsativos* bajan notablemente, manteniéndose por debajo de los valores obtenidos en datos *normativos* para todas las condiciones de *no invarianza*. Veremos más adelante, en los resultados correspondientes a ANOVA, el efecto significativo de estas variables independientes sobre el índice de ajuste *CFI*.

Tabla 7.
Estadísticos descriptivos para CFI en cada una de las condiciones del estudio

J	P	N	Invariante				No invariante			
			Normativo		Ipsatizado		Normativo		Ipsatizado	
			M	SD	M	SD	M	SD	M	SD
CF	P=9	100	.993	.011	.994	.010	.910	.031	.948	.030
		300	.997	.004	.997	.005	.906	.017	.942	.016
		500	.999	.002	.998	.003	.907	.012	.944	.011
		1000	.999	.001	.999	.001	.907	.008	.943	.008
	P=12	100	.989	.012	.989	.014	.893	.030	.920	.029
		300	.998	.003	.998	.004	.899	.014	.923	.015
		500	.999	.002	.999	.002	.901	.011	.923	.011
		1000	.999	.001	.999	.001	.901	.007	.925	.007
	P=18	100	.985	.013	.980	.018	.882	.026	.905	.028
		300	.997	.004	.996	.005	.895	.012	.918	.014
		500	.999	.002	.998	.003	.895	.010	.920	.010
		1000	.999	.001	.999	.001	.895	.006	.920	.007
MC	P=9	100	.992	.011	.992	.012	.893	.032	.872	.036
		300	.997	.004	.997	.005	.893	.018	.870	.022
		500	.999	.002	.998	.003	.893	.014	.870	.016
		1000	.999	.001	.999	.001	.893	.009	.870	.011
	P=12	100	.988	.013	.987	.015	.879	.033	.864	.038
		300	.998	.003	.997	.004	.886	.015	.871	.017
		500	.999	.002	.999	.002	.888	.011	.873	.012
		1000	.999	.001	.999	.001	.888	.008	.873	.009
	P=18	100	.985	.013	.980	.018	.870	.026	.819	.032
		300	.997	.004	.996	.005	.884	.013	.837	.016
		500	.999	.002	.998	.003	.883	.010	.838	.012
		1000	.999	.001	.999	.001	.883	.006	.838	.008

Nota. CF = Nivel de invarianza configural; MC = Nivel de invarianza métrico

La tabla 8 muestra la proporción de soluciones para las cuales ha sido rechazada la hipótesis de ajuste de *invarianza*, utilizando como criterio valores de *CFI* inferiores a .95. Estos resultados replican los resultados anteriores. Mientras que en condiciones de *invarianza* la tasa de Error Tipo I se mantiene dentro de valores

aceptables, *CFI* pierde potencia para detectar la falta de *invarianza configural* con datos *ipsatizados*, particularmente en la condición $P=9$ y con muestras pequeñas. Por otro lado, una vez ajustado el nivel de *invarianza métrica* la proporción de rechazo de ajuste es similar entre los datos *normativos* e *ipsatizados*.

Tabla 8.
Proporción de soluciones con ajuste aceptable según *CFI* ($>.95$) en cada una de las condiciones

<i>P</i>	<i>N</i>	<i>Invariante</i>				<i>No invariante</i>			
		<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>	
		<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>
$P=9$	100	.004	.004	.004	.016	.915	.970	.488	1.000
	300	.000	.000	.000	.000	1.000	1.000	.692	1.000
	500	.000	.000	.000	.000	1.000	1.000	.719	1.000
	1000	.000	.000	.000	.000	1.000	1.000	.820	1.000
$P=12$	100	.000	.008	.019	.023	.978	.994	.838	.990
	300	.000	.000	.000	.000	1.000	1.000	.965	1.000
	500	.000	.000	.000	.000	1.000	1.000	.995	1.000
	1000	.000	.000	.000	.000	1.000	1.000	1.000	1.000
$P=18$	100	.004	.004	.066	.076	1.000	1.000	.954	1.000
	300	.000	.000	.000	.000	1.000	1.000	.994	1.000
	500	.000	.000	.000	.000	1.000	1.000	.998	1.000
	1000	.000	.000	.000	.000	1.000	1.000	1.000	1.000

Nota. *CF* = Nivel de invarianza configural; *MC* = Nivel de invarianza métrico

En cuanto al estadístico *RMSEA*, los resultados son similares a los obtenidos con *CFI* en cuanto a las diferencias a nivel global entre ambos tipos de datos. En la tabla 9 podemos ver que, mientras que bajo la condición de *invarianza*, *RMSEA* alcanza valores muy similares en ambos tipos de datos, en la condición de *no invarianza*, *RMSEA* muestra dificultades para detectar la falta de *invarianza*. A diferencia de *CFI* que presenta dificultades en la condición $P=9$, para el caso de

RMSEA, los valores más bajos se obtienen en la condición $P=18$, los cuales se encuentran en torno a valores que indican un ajuste aceptable (.06).

Tabla 9.
Estadísticos descriptivos para RMSEA en cada una de las condiciones del estudio

J	P	N	Invariante				No invariante			
			Normativo		Ipsatizado		Normativo		Ipsatizado	
			M	SD	M	SD	M	SD	M	SD
CF	P=9	100	.019	.022	.018	.022	.096	.018	.073	.026
		300	.011	.013	.012	.015	.099	.009	.080	.012
		500	.007	.010	.008	.011	.099	.007	.080	.009
		1000	.006	.007	.007	.008	.099	.004	.081	.006
	P=12	100	.024	.020	.023	.022	.092	.014	.079	.017
		300	.009	.011	.009	.011	.090	.007	.080	.008
		500	.007	.008	.007	.009	.089	.005	.080	.006
		1000	.004	.006	.005	.006	.089	.003	.079	.004
	P=18	100	.026	.016	.025	.017	.082	.009	.064	.011
		300	.009	.009	.009	.010	.077	.005	.060	.006
		500	.006	.007	.006	.007	.077	.004	.060	.004
		1000	.004	.005	.004	.005	.077	.002	.060	.003
MC	P=9	100	.019	.021	.018	.022	.097	.015	.105	.017
		300	.010	.012	.012	.013	.098	.008	.108	.010
		500	.007	.009	.008	.011	.098	.007	.108	.007
		1000	.006	.007	.006	.008	.098	.004	.108	.005
	P=12	100	.023	.020	.023	.021	.093	.014	.097	.015
		300	.009	.010	.009	.011	.091	.006	.096	.007
		500	.007	.008	.007	.008	.090	.004	.096	.005
		1000	.004	.005	.005	.006	.090	.003	.096	.004
	P=18	100	.025	.016	.025	.017	.083	.009	.086	.009
		300	.009	.009	.009	.009	.079	.004	.082	.005
		500	.006	.007	.006	.007	.079	.003	.082	.003
		1000	.004	.005	.004	.005	.079	.002	.082	.002

Nota. CF = Nivel de invarianza configural; MC = Nivel de invarianza métrico

La proporción de rechazo de la hipótesis de ajuste corrobora los resultados anteriores (ver tabla 10). Podemos ver que en la condición de *no invarianza* con datos ipsatizados, el índice *RMSEA* posee tasas de Error Tipo II dentro de los valores aceptables, a excepción de las condiciones $P=18$, en las cuales existe una pérdida de potencia importantes, en torno a .50, a diferencia del índice *CFI* el cual presentaba problemas de sensibilidad en las condiciones de $P=9$.

Tabla 10
Proporción de soluciones con desajuste según *RMSEA* ($>.06$) en cada una de las condiciones del estudio

<i>P</i>	<i>N</i>	<i>Invariante</i>				<i>No invariante</i>			
		<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>	
		<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>	<i>CF</i>	<i>MT</i>
$P=9$	100	.040	.028	.056	.044	.965	.985	.721	1.000
	300	.000	.000	.000	.000	1.000	1.000	.939	1.000
	500	.000	.000	.000	.000	1.000	1.000	.989	1.000
	1000	.000	.000	.000	.000	1.000	1.000	1.000	1.000
$P=12$	100	.034	.015	.042	.030	.975	.990	.873	.987
	300	.000	.000	.000	.000	1.000	1.000	.988	1.000
	500	.000	.000	.000	.000	1.000	1.000	1.000	1.000
	1000	.000	.000	.000	.000	1.000	1.000	1.000	1.000
$P=18$	100	.002	.002	.002	.002	.986	.996	.681	.998
	300	.000	.000	.000	.000	1.000	1.000	.528	1.000
	500	.000	.000	.000	.000	1.000	1.000	.486	1.000
	1000	.000	.000	.000	.000	1.000	1.000	.508	1.000

Nota. *CF* = Nivel de invarianza configural; *MC* = Nivel de invarianza métrico

En cuanto al índice *SRMR*, los resultados descriptivos se muestran en la tabla 11. Como puede verse, para la condición de *invarianza* a nivel configural (*CF*) la media de *SRMR* alcanza valores más bajos y con menor dispersión en datos

normativos que con datos *ipsatizados*, cuyas medias y desviaciones son más elevadas.

A pesar de ello todos los valores se encuentran dentro de los límites esperados para ambos tipos de datos.

Tabla 11
Estadísticos descriptivos para SRMR en cada una de las condiciones del estudio

<i>J</i>	<i>P</i>	<i>N</i>	<i>Invariante</i>				<i>No invariante</i>			
			<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>	
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>CF</i>	<i>P=9</i>	<i>100</i>	.051	.007	.066	.015	.087	.011	.093	.015
		<i>300</i>	.029	.004	.040	.014	.076	.007	.079	.010
		<i>500</i>	.023	.003	.030	.013	.074	.006	.076	.009
		<i>1000</i>	.016	.002	.020	.009	.073	.004	.073	.006
	<i>P=12</i>	<i>100</i>	.056	.006	.075	.012	.094	.011	.108	.013
		<i>300</i>	.032	.003	.047	.015	.083	.006	.095	.010
		<i>500</i>	.025	.002	.036	.015	.080	.005	.091	.008
		<i>1000</i>	.017	.002	.024	.012	.078	.004	.087	.006
	<i>P=18</i>	<i>100</i>	.059	.005	.073	.005	.099	.009	.107	.009
		<i>300</i>	.034	.002	.042	.003	.086	.005	.091	.006
		<i>500</i>	.026	.002	.032	.002	.084	.004	.087	.005
		<i>1000</i>	.019	.001	.023	.002	.082	.003	.084	.004
<i>MC</i>	<i>P=9</i>	<i>100</i>	.081	.016	.081	.015	.119	.022	.126	.017
		<i>300</i>	.046	.009	.048	.014	.099	.015	.111	.011
		<i>500</i>	.036	.008	.038	.016	.092	.012	.107	.008
		<i>1000</i>	.025	.005	.030	.019	.090	.008	.105	.006
	<i>P=12</i>	<i>100</i>	.083	.015	.087	.015	.122	.022	.132	.018
		<i>300</i>	.048	.009	.052	.015	.100	.013	.113	.010
		<i>500</i>	.037	.006	.038	.011	.095	.011	.109	.008
		<i>1000</i>	.026	.004	.027	.008	.092	.008	.106	.006
	<i>P=18</i>	<i>100</i>	.085	.013	.087	.008	.125	.019	.134	.011
		<i>300</i>	.049	.008	.050	.004	.101	.012	.113	.006
		<i>500</i>	.038	.006	.038	.005	.097	.010	.109	.005
		<i>1000</i>	.027	.004	.027	.004	.093	.007	.106	.004

Nota. CF = Nivel de invarianza configural; MT = Nivel de invarianza métrico

Por otro lado, en la condición de *no invarianza* el comportamiento de *SRMR* también difiere en comparación con el de los índices *CFI* y *RMSEA*. Para todas las condiciones de *no invarianza*, *SRMR* alcanza valores más altos en datos *ipsatizados* que en datos *normativos*. Este hallazgo difiere del resto de los índices de ajuste. Mientras *CFI* y *RMSEA* tienen dificultades para detectar la falta de invarianza en datos *ipsatizados*, *SRMR* parece ser más sensible a la falta de invarianza y menos sensible a las variaciones muestrales con datos *ipsatizados* que con datos *normativos*.

La proporción de rechazo del ajuste también corrobora los resultados anteriores (ver tabla 12). En todas las condiciones de *no invarianza* el modelo *ipsatizado* logra proporciones de rechazo mayores que el modelo *normativo*, aunque en algunas condiciones de $P=9$ la pérdida de potencia es notable: la tasa de rechazo alcanza valores muy bajos, en algunos casos no superando el 10% (por ejemplo $N=1000$). El mismo fenómeno ocurre en la condición $P=12$ para datos *normativos*, no así para los datos *ipsatizados*, en los cuales la proporción de rechazo no baja del 90% para todos los tamaños muestrales.

Estos resultados muestran que a diferencia de *CFI* y *RMSEA*, el índice *SRMR* es más sensible a la falta de invarianza configural con datos *ipsatizados*, particularmente cuando el número de variables en el modelo es mayor o igual a 12.

Tabla 12.
Proporción de rechazo del ajuste según SRMR ($>.08$) en cada una de las condiciones del estudio

P	N	Invariante				No invariante			
		Normativo		Ipsatizado		Normativo		Ipsatizado	
		CF	MT	CF	MT	CF	MT	CF	MT
P=9	100	.000	.448	.008	.321	.731	.995	.756	1.000
	300	.000	.005	.000	.000	.294	.928	.400	.998
	500	.000	.000	.000	.000	.162	.862	.208	1.000
	1000	.000	.000	.000	.000	.037	.900	.075	1.000
P=12	100	.000	.498	.046	.540	.921	1.000	.994	1.000
	300	.000	.006	.000	.000	.658	.948	.916	1.000
	500	.000	.000	.000	.000	.507	.923	.930	1.000
	1000	.000	.000	.000	.000	.275	.946	.913	1.000
P=18	100	.000	.611	.096	.816	.996	1.000	.998	1.000
	300	.000	.002	.000	.000	.888	.988	.962	1.000
	500	.000	.000	.000	.000	.830	.966	.886	1.000
	1000	.000	.000	.000	.000	.772	.982	.862	1.000

Nota. CF = Nivel de invarianza configural; MC = Nivel de invarianza métrico

Para determinar el efecto de las condiciones de estudio hemos realizado un ANOVA factorial para cada uno de los índices de ajuste estudiados. Los resultados se muestran en la tabla 13. Para el caso de CFI los mayores efectos encontrados corresponden a I ($\eta^2 = .948$) y J ($\eta^2 = .848$) para los efectos principales, y a $D*J$ ($\eta^2 = .845$) y $J*I$ ($\eta^2 = .851$) para los efectos de doble interacción, todos ellos con tamaños grandes (en torno a .80). Adicionalmente, uno de los resultados más interesantes corresponde al efecto de triple interacción $D*J*I$ ($\eta^2 = .849$), el cual se representa gráficamente en la figura 9.

Tabla 13.
ANOVA factorial para la diferencia en los índices de ajuste CFI, RMSEA y SRMR

	<i>gl</i>	<i>CFI</i>			<i>RMSEA</i>			<i>SRMR</i>		
		<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
<i>D</i>	1	18.917	.000	.002	1960.886	.000	.162	7472.118	.000	.424
<i>J</i>	1	56651.072	.000	.848	13668.009	.000	.574	46413.634	.000	.820
<i>I</i>	1	1867755.639	.000	.948	165758.993	.000	.942	255130.219	.000	.962
<i>P</i>	2	1110.710	.000	.179	1042.028	.000	.170	1309.870	.000	.205
<i>N</i>	3	267.443	.000	.073	415.845	.000	.109	17837.834	.000	.840
<i>D * I</i>	1	20.092	.000	.002	2119.383	.000	.173	1594.988	.000	.136
<i>D * P</i>	2	1817.416	.000	.263	155.684	.000	.030	187.308	.000	.036
<i>D * N</i>	3	4.323	.005	.001	58.946	.000	.017	11.325	.000	.003
<i>D * J</i>	1	58184.407	.000	.851	29822.234	.000	.746	.552	.470	.000
<i>J * I</i>	2	55233.373	.000	.845	15406.039	.000	.603	3841.294	.000	.274
<i>J * P</i>	2	616.915	.000	.108	231.793	.000	.044	460.959	.000	.083
<i>J * N</i>	3	20.682	.000	.006	7.219	.000	.002	964.481	.000	.222
<i>I * P</i>	2	693.803	.000	.120	10921.497	.000	.177	75.503	.000	.015
<i>I * N</i>	3	17.931	.000	.005	338.309	.000	.091	2393.420	.000	.414
<i>P * N</i>	6	59.002	.000	.034	35.084	.000	.020	41.451	.000	.024
<i>D * I * P</i>	2	1317.445	.000	.206	142.742	.000	.027	93.381	.000	.018
<i>D * I * N</i>	3	4.054	.007	.001	3.712	.000	.001	64.740	.000	.019
<i>D * P * N</i>	6	25.529	.000	.015	6.490	.000	.004	16.616	.000	.010
<i>J * I * P</i>	2	625.093	.000	.110	235.382	.000	.004	196.945	.000	.037
<i>J * I * N</i>	3	7.945	.000	.002	14.938	.000	.004	40.805	.000	.012
<i>J * P * N</i>	6	.665	.678	.000	9.468	.000	.006	1.917	.074	.001
<i>D * J * I</i>	1	57195.620	.000	.849	29153.297	.000	.742	3568.661	.000	.026
<i>D * J * P</i>	2	1611.714	.000	.241	846.305	.000	.143	89.488	.000	.017
<i>D * J * N</i>	3	9.843	.000	.003	12.534	.000	.004	265.181	.000	.073
<i>I * P * N</i>	6	4.015	.001	.002	.653	.000	.000	3.704	.001	.002

Nota. *D* = Tipo de datos; *J* = Nivel de invarianza; *I* = Tipo de invarianza; *P* = N° de variable y *N* = Tamaño de la muestra

La figura 9.a muestra el comportamiento de *CFI* en la condición de *invarianza*. Como puede observarse los resultados son similares en datos *normativos* e *ipsatizados*, tanto a nivel configural como métrico. Por el contrario, en condiciones de *no invarianza* (véase figura 9.b), el índice *CFI* alcanza valores significativamente mayores en datos *ipsatizados* a nivel configural, mostrando dificultades para detectar la falta de *invarianza*. Por otro lado, una vez ajustado el nivel métrico, los valores de *CFI* disminuyen quedando significativamente por debajo de los valores obtenidos por los datos *normativos*.

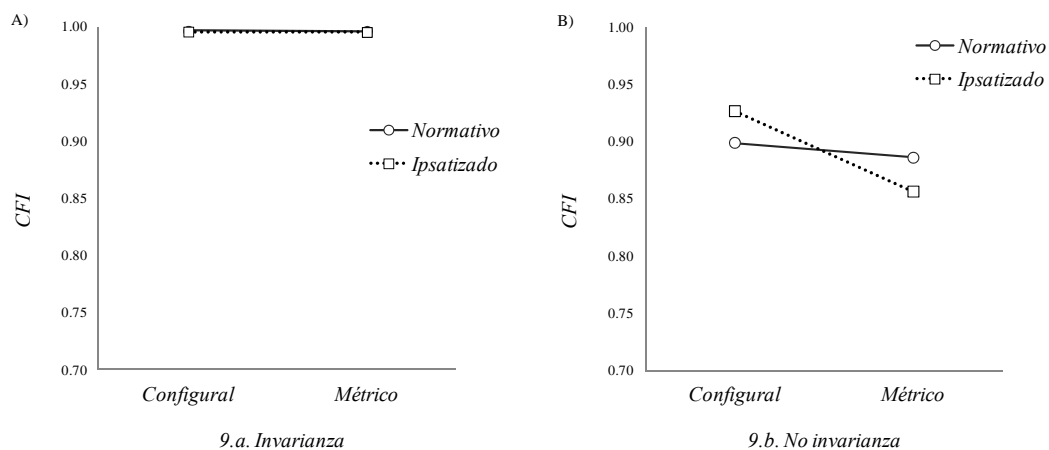


Figura 9. Efecto de triple interacción $D*I*P$ para *CFI*

El resto de los efectos de triple interacción poseen un tamaño del efecto bajo. Estos corresponden a los efectos $D*J*P$ ($\eta^2 = .241$) y $D*I*P$ ($\eta^2 = .206$). Para el caso del efecto de la interacción $D*J*P$ (ver figura 10), podemos ver que a nivel configural, *CFI* mantiene valores más bajos con datos *normativos* que con datos *ipsatizados*, aunque estos valores van disminuyendo en la medida que aumenta la cantidad de ítems en el modelo, reduciéndose las diferencias entre datos *normativos* e

ipsatizados (figura 10.a). En cuanto al nivel de invarianza métrico, son los datos *ipsatizados* los que obtienen valores menores en *CFI* en comparación con los datos *normativos*. Mientras que los valores de *CFI* con datos *normativos* no se ven modificados frente a la cantidad de ítems, en el modelo *ipsatizado* el valor de *CFI* disminuye cuando el n° de ítems en el modelo aumenta.

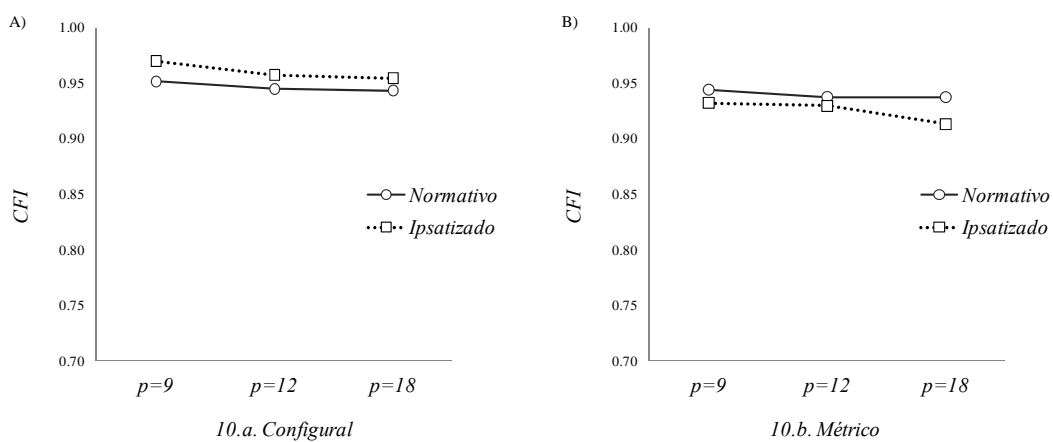


Figura 10 . Efecto de triple interacción $D*J*P$ para *CFI*

Un fenómeno similar sucede con la interacción $D*I*P$ (ver figura 11). Frente a la condición de *invarianza*, *CFI* posee el mismo comportamiento con datos *ipsatizados* que con datos *normativos*, manteniendo valores similares a medida que la complejidad del modelo aumenta. Por el contrario, en la condición de *no invarianza*, los valores de *CFI* para datos *ipsatizados* disminuyen progresivamente a medida que aumenta el número de ítems, hasta alcanzar valores ligeramente por debajo del modelo *normativo*.

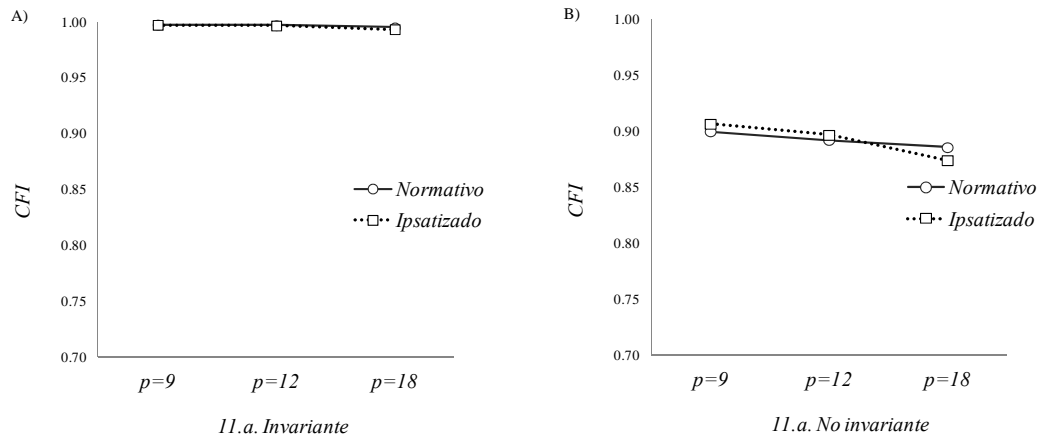


Figura 11. Efecto de triple interacción $D*I*P$ para CFI

Globalmente, los resultados en cuanto al índice *RMSEA* son similares a los obtenidos con el índice *CFI*. Para *RMSEA* los efectos principales más importantes corresponden a los efectos de *I*: condición de invarianza ($\eta^2 = .942$) y *J*: nivel de invarianza ($\eta^2 = .574$), y a los efectos de doble interacción $D*J$ ($\eta^2 = .746$) y $J*I$ ($\eta^2 = .603$). En cuanto a los efectos de triple interacción el más importante corresponde al efecto $D*J*I$ ($\eta^2 = .742$). La figura 12 muestra en más detalle este efecto.

Al igual que *CFI*, el comportamiento del índice *RMSEA* es similar en el modelo *normativo* y en el *ipsatizado*, frente la condición de invarianza (12.a). En cambio, en la condición de *no invarianza* (12.b) el modelo *ipsatizado* obtiene valores menores cuando se contrasta el nivel de invarianza *configural*, mientras que cuando se contrasta el nivel de invarianza *métrico*, el modelo *ipsatizado* alcanza valores superiores en *RMSEA* que en el modelo *normativo*.

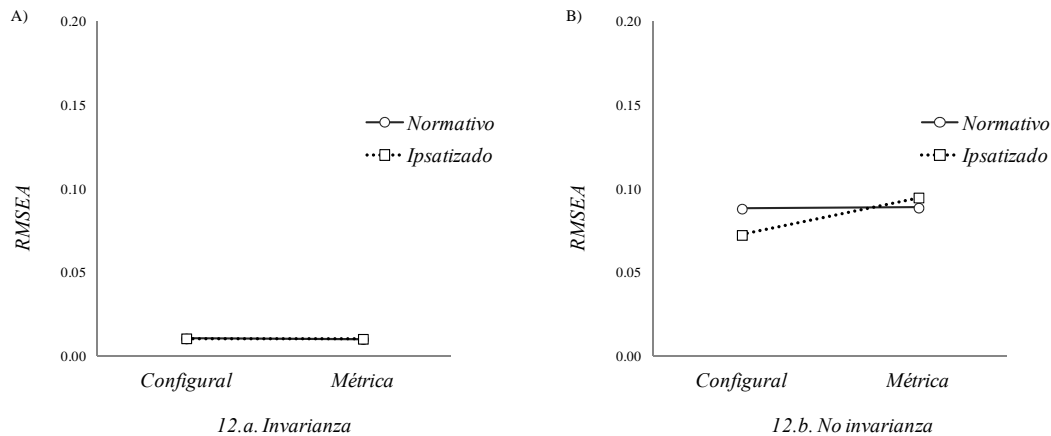


Figura 12. Efecto de triple interacción $D*J*I$ para RMSEA

En cuanto al índice *SRMR*, en la línea de lo encontrado en los análisis descriptivos, los resultados difieren en relación a lo hallado con *CFI* y *RMSEA*. Junto con los efectos principales de *I* ($\eta^2 = .962$) y *J* ($\eta^2 = .820$), existe un efecto grande de *N* ($\eta^2 = .840$) y mediano de *D* ($\eta^2 = .424$), los cuales no están presentes en los índices *CFI* y *RMSEA*. En cuanto a los efectos de la doble interacción, los más importantes corresponden a $N*I$ ($\eta^2 = .414$), con un tamaño del efecto medio, y $J*N$ ($\eta^2 = .222$) y $J*I$ ($\eta^2 = .274$), con un tamaño del efecto bajo.

Los detalles de los efectos de doble interacción se muestran en la figura 13. Nótese que, a pesar de haber un efecto principal del *tipo de datos* (*D*), este no interactúa con ningún otro efecto principal, a diferencia de *CFI* y *RMSEA*, los cuales se ven afectados por la interacción de tipo de datos (*D*), el tipo de invarianza (*I*) y el nivel de invarianza (*J*). Por tanto el efecto del tipo de datos sobre *SRMR* es independiente del resto de las condiciones, siendo sus valores más altos en datos *normativos* que en *ipsatizados*, independientemente de la presencia de otros factores.

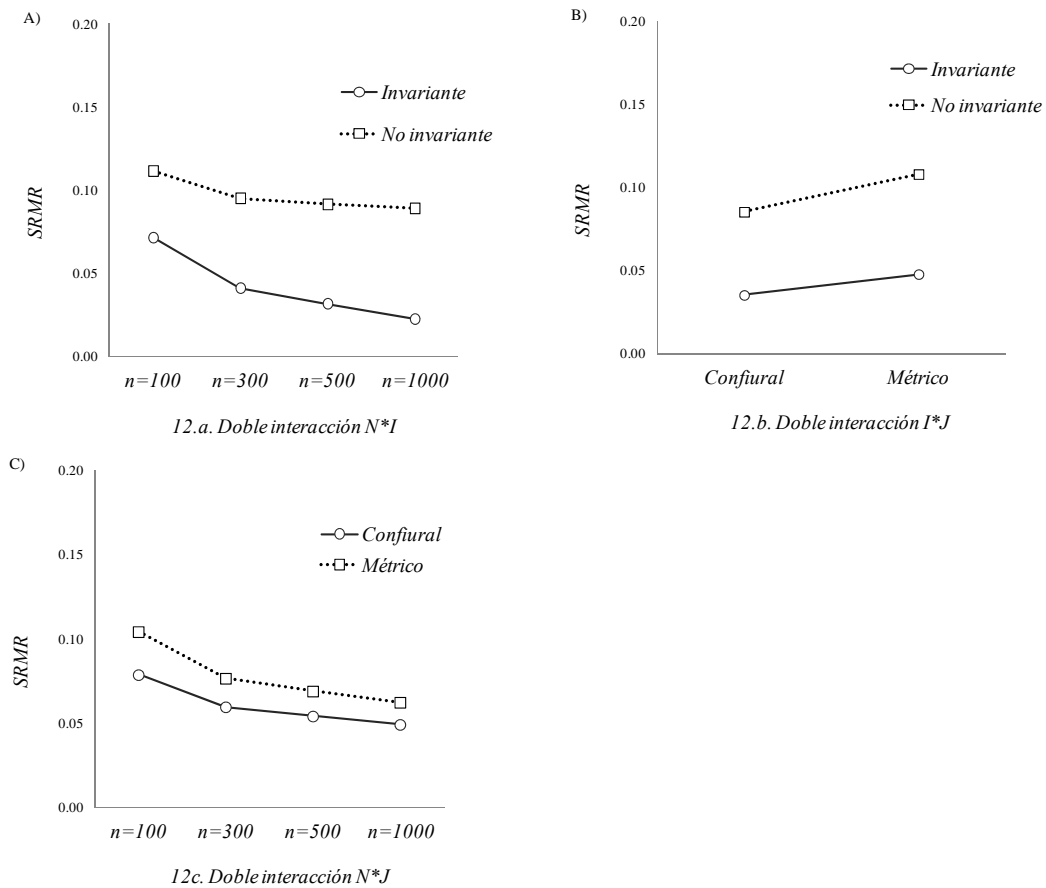


Figura 13. Efecto de doble interacción para SRMR

En cuanto a los efectos de la doble interacción, el más importante corresponde al efecto $N*I$, el cual se presenta en detalle en la figura 13.a. Podemos ver que la condición de *invarianza* obtiene una media en *SRMR* significativamente inferior a la obtenida por la condición de *no invarianza*. Esta diferencia se hace aún mayor en la medida que aumenta el tamaño de la muestra, debido a que, en condición de *invarianza*, la media de *SRMR* disminuye al aumentar N , mientras que para la condición *no invarianza*, *SRMR* disminuye en menor medida.

En cuanto al resto de los efectos de doble interacción, tanto $I*J$ ($\eta^2=.222$) como $N*J$ ($\eta^2=.274$) poseen efectos más pequeños. El gráfico 13.b muestra que, tanto en grupos *invariantes* como *no invariantes*, el índice *SRMR* es sensible al nivel de invarianza, aunque los valores mayores se obtienen cuando los grupos son no invariantes. El gráfico 13.c muestra la interacción $N*J$, y pone de manifiesto la existencia de diferencias en *SRMR* entre los niveles de invarianza *configural* y *métrico*, independientemente si los grupos son invariantes o no. En la medida en que aumenta el tamaño de la muestra disminuyen los valores de *SRMR*, así como su diferencia entre los dos niveles de invarianza.

5.1.5. Conclusiones

En general, podemos afirmar que los resultados obtenidos han sido los esperados para los modelos correctamente especificados. De forma similar al estudio de Ximénez y Calderón (2012), pero en este caso para la condición de que los grupos son *invariantes*, el funcionamiento de los índices de ajuste *CFI* y *RMSEA* con datos *ipsatizados* es similar que con datos *normativos*. En cambio, cuando los grupos son *no invariantes*, o lo que es lo mismo, bajo errores de especificación, tanto el índice *CFI* como el índice *RMSEA* tienen algunas dificultades para detectar la *no invarianza configural*. En el caso de *CFI*, estas dificultades se presentan cuando el modelo tiene pocas variables ($P=9$) y el tamaño de la muestra es reducido ($N<300$). En cambio, el índice *RMSEA* se torna conservador cuanto mayor es el número de variables en el modelo ($P=18$), aunque esta variable tiene un efecto pequeño. Estas dificultades se corrigen cuando se restringen las saturaciones factoriales a ser iguales entre los

grupos (bajo el supuesto de *invarianza métrica*). En tales casos, el comportamiento de los índices *CFI* y *RMSEA* es similar en datos *normativos* y *ipsatizados* frente a diferencias en la estructura factorial entre los grupos.

Adicionalmente, existe una pérdida de la sensibilidad para detectar la falta de invarianza en los modelos ajustados con datos *ipsatizados*, en algunas de las condiciones estudiadas. No obstante, cabe señalar que las dificultades de los índices *CFI* y *RMSEA* no aparecen en el estadístico χ^2 , el cual no presenta una pérdida significativa de la potencia estadística. Por tanto, a nivel configural, el contraste de la hipótesis de ajuste a través del estadístico χ^2 , así como la interpretación de los índices *CFI* y *RMSEA* se puede llevar a cabo del mismo modo que con datos *normativos*, aunque con las excepciones de $P=9$, para el caso de *CFI* y $P=18$ para el caso de *RMSEA*. A pesar de ello, de igual forma recomendamos a los investigadores ser cautelosos en la utilización de estos índices, debido a que los resultados del ANOVA mostraron que sus valores se ven bastante alterados en condiciones de no invarianza con datos *ipsatizados*.

Por otro lado, el índice *SRMR*, que no había sido estudiando con anterioridad, no presenta un funcionamiento similar al de los anteriores índices de bondad de ajuste. En todas las condiciones, con ambos tipos de datos y tanto en condiciones de *invarianza* como de *no invarianza*, *SRMR* se ve afectado por el tamaño de la muestra, tornándose conservador en la medida que la muestra aumenta.

En condiciones de *invarianza* existen elevadas tasas de error tipo I en muestras pequeñas ($N=100$), tanto con datos *normativos* como con datos *ipsatizados*. Esto lleva a *SRMR* a rechazar la hipótesis de la igualdad entre grupos que se asumen

equivalentes a nivel configural cuando las muestras se encuentran en torno a 100 observaciones. En cambio, frente a condiciones de *no invarianza*, no existen las diferencias encontradas en *CFI* y *RMSEA* entre datos *normativos* e *ipsatizados*. Para los datos *ipsatizados*, *SRMR* alcanza una potencia dentro de los valores aceptables, a excepción de los modelos con un menor número de variables ($P=9$), para los cuales existe una pérdida sustancial de la potencia. Por tanto, *SRMR* se convierte en una buena alternativa cuando se sospecha de la existencia de no invarianza configural, siempre y cuando el modelo tenga 12 o más variables y el tamaño de la muestra sea el adecuado.

5.2. ESTUDIO 2: SENSIBILIDAD DE LOS ÍNDICES DE AJUSTE A LA INVARIANZA MÉTRICA EN EL AFCM

El presente estudio se plantea como una extensión del estudio anterior, y tiene como propósito complementar el estudio del funcionamiento del AFCM con variables *ipsatizadas*. Esta vez, el objetivo es evaluar la sensibilidad de los índices de bondad de ajuste a la *no invarianza* a nivel métrico en modelos con la misma estructura factorial. A diferencia del estudio 1, en el cual hemos introducido la condición de no invarianza a nivel configural a través del error de especificación por alteración del número de factores, en el presente estudio hemos introducido la *no invarianza* a través de las diferencias en las saturaciones factoriales entre los grupos con la misma estructura factorial.

5.2.1. Diseño

Al igual que en el estudio 1, el diseño ha considerado dos grupos. Para todas las condiciones hemos generado una estructura de tres factores. La tabla 14 muestra las variables consideradas en el estudio. Las *variables independientes* son:

- 1) Patrón Invarianza (I): Este factor posee tres niveles, uno de *invarianza* y dos de *no invarianza*. Por un lado, [1] la condición de *invarianza*, en la cual tanto la estructura factorial como las saturaciones factoriales son iguales en ambos grupos. Al igual que en el estudio anterior, los valores teóricos de los parámetros (ver anexo 5) han sido los utilizados por Chan y Bentler (1996) y Ximénez y Calderón (2012). Estos valores se han utilizado como referencia para generar el

resto de los modelos. Por otro lado, [2] la condición de *no invarianza homogénea*, donde todos los valores de las saturaciones factoriales fueron más bajos en el grupo 2, y [3] la condición de *no invarianza mixta*, donde los parámetros fueron modificados en ambos grupos de tal manera que la mitad de los saturaciones fueron más bajas en el grupo 1 y la otra mitad más bajas en el grupo 2.

- 2) Tipo de datos (D): al igual que en el estudio 1 hemos implementado los análisis con datos *normativos (pre-ipsativos)* y posteriormente con sus correspondientes datos *ipsatizados*.
- 3) Nº de variables (P) y tamaño de la muestra (N). Del mismo modo que en el estudio anterior, hemos evaluado el efecto del número de variables en el modelo: 9, 12 y 18 ítems, y el tamaño de la muestra: 100, 300, 500 y 1000 observaciones.

En cuanto a las ***variables dependientes***, a diferencia del estudio 1, en el que habíamos incorporado el *nivel de invarianza* como un factor, en el presente estudio lo hemos incorporado a través del análisis de la diferencia en los índices de bondad de ajuste entre los niveles configural y métrico (asumiendo la invarianza a nivel configural). Para el caso de la hipótesis de ajuste, hemos analizado el test de diferencia de χ^2 ; y para el resto de índices de ajuste hemos utilizado la diferencia de los valores entre cada uno de los niveles de invarianza. Hemos utilizando los criterios propuestos por Chen (2007) para determinar una pérdida de ajuste significativa entre ambos niveles: $CFI > .010$; $RMSEA > .010$ y $SRMR > .025$.

Tabla 14.
Variables consideradas en el estudio 2

Código	Variable	Niveles
<i>Variables independientes</i>		
<i>I</i>	Patrón de invarianza	Invarianza No invariante homogénea No invariante mixta
<i>D</i>	Tipo de datos	Normativos Ipsatizados
<i>P</i>	N° de indicadores	9 12 18
<i>N</i>	Tamaño de la muestra	100 300 500 1000
<i>Variables dependientes</i>		
Δ_{PER}	Proporción empírica de rechazo de la hipótesis de ajuste del nivel métrico	
Δ_{CFI}	Comparative Fit Index	(>.010)
Δ_{RMSEA}	Root-Mean Square Error of Approximation	(>.010)
Δ_{SRMR}	Standardized Root Mean Square Residual	(>.025)

5.2.2. Procedimiento

El primer paso ha sido diseñar las estructuras de base a partir de las cuales han sido generados el resto de los modelos. Estas estructuras corresponden a tres modelos de tres factores, con 9, 12 y 18 variables cada uno. Los valores de los parámetros han sido definidos para coincidir con los valores estandarizados de .707 para las saturaciones factoriales y .50 para los errores de medida (para cumplir $\lambda^2 + \varepsilon = 1$). Estas estructuras han sido utilizadas como estructura de referencia para el resto de los modelos (ver modelo 1 del anexo 5).

Para la condición de igualdad entre los grupos (*patrón invariante*), hemos simulado los grupos 1 y 2 a partir de los valores de los parámetros del modelo 1. Para el caso de la condición de *invarianza homogénea*, sólo han sido modificados los parámetros del grupo 2 para ser menores que los del grupo 1. Todas las saturaciones han sido modificadas para que coincidan con los valores estandarizados de .50 para las saturaciones factoriales y .75 para los errores de medida. Es así que para esta condición, el grupo 1 quedó constituido por los valores del modelo 1, y el grupo 2 por los valores del modelo 2 (ver anexo 5).

Finalmente, para la condición de *no invarianza mixta*, los valores de los parámetros base (modelo 1) fueron modificados de manera alterna, de forma que la mitad de las saturaciones son menores en el grupo 1, y la otra mitad son menores en el grupo 2. Por tanto, las saturaciones con valores altos corresponden a las del modelo 1 ($\lambda = .707$ y $\varepsilon = .50$) y con valores bajos al modelo 2 ($\lambda = .50$ y $\varepsilon = .75$).

Para cada condición se simularon 500 réplicas con distribución normal multivariante con media $\mu_y = 0$ y matriz de covarianzas $\Sigma = \Lambda_y \Phi_y \Lambda_y' + \Psi_y$. Una vez generados los datos *normativos*, estos han sido transformados a sus correspondientes datos *ipsatizados* a través de la ecuación [1.10]. Ambos conjuntos de datos se analizaron mediante el AFCM; para el caso de los datos *normativos* ha sido aplicado un AFCM convencional, y para el caso de los datos *ipsatizados*, el AFCM con el método *DE* de Cheung y Chan (2004). Tanto para la generación de los datos como para el ajuste de los AFCM hemos utilizado el programa EQS 6.1 de Bentler (2006).

5.2.3. Análisis de datos

Al igual que en el estudio anterior, hemos realizado un análisis descriptivo de las variables dependientes en cada una de las condiciones del estudio. Adicionalmente, hemos calculado la proporción de soluciones de rechazo del ajuste según los criterios sugeridos por Chen (2007), para el análisis de invarianza métrica (i.e., $CFI > .010$; $RMSEA > .010$ y $SRMR > .025$). Para examinar el efecto de las variables independientes, hemos llevado a cabo un ANOVA factorial, y utilizado el estadístico η^2 para determinar la significación práctica de los efectos que resultaron estadísticamente significativos. Como puede verse en la tabla 14, el modelo factorial contempla cuatro variables independientes: una variable intra-sujeto *D*: el *Tipo de datos (normativo vs. ipsatizado)* y tres variables inter-sujeto: *I*: *Patrón de invarianza (invariante, no invariante homogéneo y no invariante mixto)*; *P*: *Nº de variables (9, 12 y 18)* y *N*: *Tamaño de la muestra (100, 300, 500 y 1000)*. Por tanto, el diseño del estudio queda compuesto por $(D=2) \times (I=3) \times (P=3) \times (N=4) = 72$ condiciones.

Al igual que en el estudio anterior, el meta-modelo incorpora tanto los efectos principales, como los efectos de doble y triple interacción. Nuevamente, hemos decidido analizar las triples interacciones desde un punto de vista gráfico, debido al interés de alguno de estos efectos.

5.2.4. Resultados

En cuanto a las soluciones inadecuadas, las tablas 15 y 16 muestran las proporciones de Casos Heywood y soluciones no convergentes en cada una de las condiciones de estudio. En cuanto a los caso Heywood, la mayoría de las soluciones

inadmisibles se encuentran en la condición de *no invarianza homogénea*. La proporción de casos Heywood fue mayor a nivel configural, en modelos con menos variables ($P=9$) y menor tamaño de la muestra ($N=100$). Un fenómeno similar ocurre en la condición de *no invarianza mixta*, aunque las proporciones son considerablemente menores.

En cuanto a las soluciones no convergentes, sólo un .22% del total de las soluciones no lograron la convergencia, las cuales corresponden a soluciones ajustadas con datos *ipsatizados*. Ninguna de las condiciones posee soluciones no convergentes que superen el 8%.

Tabla 15.
Proporción de Casos Heywood en cada una de las condiciones del estudio

<i>P</i>	<i>N</i>	<i>Invariante</i>				<i>No invariante homogéneo</i>				<i>No invariante mixto</i>			
		<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>		<i>Normativo</i>		<i>Ipsatizado</i>	
		<i>CF</i>	<i>MC</i>	<i>CF</i>	<i>MC</i>	<i>CF</i>	<i>MC</i>	<i>CF</i>	<i>MC</i>	<i>CF</i>	<i>MC</i>	<i>CF</i>	<i>MC</i>
<i>P=9</i>	<i>100</i>	.000	.000	.410	.050	.090	.002	.892	.232	.074	.000	.888	.422
	<i>300</i>	.000	.000	.030	.002	.000	.000	.494	.010	.002	.000	.324	.070
	<i>500</i>	.000	.000	.008	.000	.000	.000	.268	.010	.000	.000	.150	.068
	<i>1000</i>	.000	.000	.004	.000	.000	.000	.060	.000	.000	.000	.104	.060
<i>P=12</i>	<i>100</i>	.000	.000	.016	.002	.010	.000	.546	.006	.004	.000	.364	.016
	<i>300</i>	.000	.000	.008	.000	.000	.000	.070	.000	.000	.000	.006	.002
	<i>500</i>	.000	.000	.002	.000	.000	.000	.024	.000	.000	.000	.002	.006
	<i>1000</i>	.000	.000	.000	.000	.000	.000	.018	.000	.000	.000	.004	.000
<i>P=18</i>	<i>100</i>	.000	.000	.006	.000	.000	.000	.318	.000	.000	.000	.046	.004
	<i>300</i>	.000	.000	.000	.000	.000	.000	.010	.000	.000	.000	.000	.000
	<i>500</i>	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.000	.000
	<i>1000</i>	.000	.000	.000	.000	.000	.000	.004	.000	.000	.000	.000	.000

Nota. *C* = Nivel de invarianza configural; *M* = Nivel de invarianza métrico

Tabla 16
 Proporción de soluciones no convergentes en cada una de las condiciones del estudio

P	N	Invariante				No inv. homogéneo				No inv. mixto			
		Normativo		Ipsatizado		Normativo		Ipsatizado		Normativo		Ipsatizado	
		CF	MC	CF	MC	CF	MC	CF	MC	CF	MC	CF	MC
P=9	100	.000	.000	.002	.000	.000	.000	.004	.002	.000	.000	.008	.004
	300	.000	.000	.000	.000	.000	.000	.006	.000	.000	.000	.006	.002
	500	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.000	.004
	1000	.000	.000	.000	.000	.000	.000	.002	.000	.000	.000	.004	.000
P=12	100	.000	.000	.000	.000	.000	.000	.008	.000	.000	.000	.002	.004
	300	.000	.000	.004	.000	.000	.000	.002	.000	.000	.000	.000	.000
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.002	.000
P=18	100	.000	.000	.000	.000	.000	.000	.006	.000	.000	.000	.000	.002
	300	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	500	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	1000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Nota. C = Nivel de invarianza configural; M = Nivel de invarianza métrico

En relación a la diferencia en el ajuste entre el modelo base y el nivel de invarianza métrico, la tabla 17 muestra la proporción de rechazo de la hipótesis de igualdad según el estadístico χ^2 entre ambos niveles. Hay tres resultados relevantes. En primer lugar, cuando los grupos son *no invariantes*, la mayoría de las condiciones poseen tasas de error tipo I dentro de los valores aceptables (en torno a .05). En segundo lugar, a excepción de la condición $P=9$ y $N=100$, todas las condiciones de *no invarianza mixta* poseen niveles de potencia aceptables, lo cual indicaría que en condiciones de *no invarianza mixta*, el AFCM de datos *ipsatizados* logra detectar de manera adecuada la diferencia en los coeficientes factoriales. Por último, cuando la diferencia en las saturaciones factoriales es homogénea, existe una pérdida importante

de la potencia, especialmente cuando el modelo posee pocas variables y el tamaño de la muestra es reducido. Estos resultados indican que la prueba χ^2 de diferencia en el ajuste tiene problemas para detectar la no invarianza métrica con datos *ipsatizados*, cuando el patrón de no invarianza es homogéneo.

Tabla 17.
Proporción de rechazo de la hipótesis igualdad en el ajuste entre el nivel configural y el nivel métrico

P	N	Invariante		No inv. homogéneo		No inv. mixto	
		Normativo	Ipsatizado	Normativo	Ipsatizado	Normativo	Ipsatizado
P=9	100	.078	.055	.489	.170	.822	.378
	300	.045	.054	.996	.446	.997	.877
	500	.052	.042	1	.770	1	.990
	1000	.038	.050	1	.977	1	1
P=12	100	.059	.092	.588	.261	.936	.748
	300	.050	.067	.998	.802	1	.998
	500	.046	.052	1	.977	1	1
	1000	.044	.040	1	1	1	1
P=18	100	.050	.085	.663	.378	1	.971
	300	.042	.060	.998	.818	1	1
	500	.056	.050	1	.976	1	1
	1000	.038	.048	1	1	1	1

La tabla 18 muestra los estadísticos descriptivos para la diferencia en *CFI* (Δ_{CFI}) entre el nivel configural y el métrico. Podemos observar que para el patrón de no invarianza mixto, los promedios de Δ_{CFI} para todas las condiciones se mantienen sobre .01. En cambio, cuando el patrón de no invarianza es homogéneo, Δ_{CFI} alcanza valores inferiores a .01 para la mayoría de las condiciones, indicando una pérdida significativa del ajuste.

Tabla 18.

Descriptivos para la diferencia en ajuste según Δ_{CFI} entre nivel configural y métrico para las distintas condiciones del estudio

P	N	Invariante				No invarianza homogénea				No invarianza mixto			
		Normativo		Ipsatizado		Normativo		Ipsatizado		Normativo		Ipsatizado	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
P=9	100	.001	.006	.002	.008	.018	.018	.005	.013	.033	.033	.011	.017
	300	.000	.002	.001	.003	.029	.012	.006	.007	.043	.016	.019	.013
	500	.000	.001	.000	.002	.031	.009	.008	.006	.048	.012	.021	.009
	1000	.000	.001	.000	.001	.032	.006	.009	.004	.048	.008	.023	.006
P=12	100	.001	.005	.001	.007	.017	.018	.007	.013	.044	.025	.025	.019
	300	.000	.002	.000	.002	.022	.010	.009	.007	.052	.014	.031	.011
	500	.000	.001	.000	.001	.024	.007	.010	.005	.052	.011	.031	.009
	1000	.000	.000	.000	.001	.024	.004	.010	.003	.053	.007	.032	.006
P=18	100	.000	.004	.001	.006	.015	.011	.010	.012	.048	.018	.045	.020
	300	.000	.001	.000	.002	.014	.006	.007	.005	.049	.010	.044	.011
	500	.000	.001	.000	.001	.015	.005	.007	.004	.050	.007	.044	.008
	1000	.000	.000	.000	.000	.016	.003	.008	.003	.052	.005	.046	.005

Nota. CF = Nivel de invarianza configural; MC = Nivel de invarianza métrico

Adicionalmente, hemos calculado la proporción de soluciones en las cuales debería ser rechazada la hipótesis de igualdad en el ajuste, entre el nivel configural y el nivel métrico, según el valor de Δ_{CFI} en cada una de las condiciones. Los resultados se presentan en la tabla 19. Estos resultados refuerzan la idea anterior. Por un lado, cuando el patrón de *no invarianza* es mixto la mayoría de las condiciones (a excepción de $N=9$ y $N=100$) mantienen niveles de potencia aceptables (en torno a .80). Por el contrario, cuando el patrón de *no invarianza* es homogéneo, el AFCM con datos *ipsatizados* obtienen tasas de error tipo II excesivamente altas, que en la mayoría de los casos sobrepasa el 50%.

Tabla 19.
Proporción de soluciones en las cuales existe una pérdida significativa del ajuste según CFI al ajustar el nivel métrico ($\Delta_{CFI} < .01$)

P	N	Invariante		No inv. homogéneo		No inv. mixto	
		Norm.	Ips.	Norm.	Ips.	Norm.	Ips.
P=9	100	.096	.113	.596	.170	.711	.333
	300	.000	.010	.924	.249	.984	.742
	500	.000	.002	.997	.327	1	.910
	1000	.000	.000	1	.334	1	.986
P=12	100	.061	.102	.566	.305	.914	.748
	300	.002	.000	.901	.415	.998	.982
	500	.000	.000	.980	.477	1	1.000
	1000	.000	.000	1	.550	1	1.000
P=18	100	.014	.068	.651	.431	.994	.968
	300	.000	.000	.774	.261	1	.998
	500	.000	.000	.854	.204	1	1.000
	1000	.000	.000	.966	.169	1	1.000

Nota. Norm. = Normativo; Ips. = Ipsatizado.

Las tablas 20 y 21 muestran los resultados de $RMSEA$ correspondientes a la diferencia en el ajuste y la proporción de rechazo de la igualdad en el ajuste, respectivamente. Podemos ver que los resultados obtenidos para Δ_{RMSEA} son similares a los obtenidos con Δ_{CFI} aunque con algunas diferencias. En primer lugar, tanto la media como la proporción de rechazo del ajuste indican que Δ_{RMSEA} tiene dificultades para detectar la no invarianza cuando el *patrón de invarianza* es homogéneo, mientras que, cuando el patrón es mixto, los valores de Δ_{RMSEA} con datos *ipsatizados* son superiores, tanto en la media como en la proporción de rechazo del ajuste, valores que incluso llegan a ser similares en muchos casos a los obtenidos con los datos

normativos. En segundo lugar, se observa que, tanto la media como la proporción de rechazo del ajuste aumentan a medida que aumenta el tamaño de la muestra y disminuye el número de variables. Esto nos indica que, a diferencia de *CFI*, el cual muestra dificultades para detectar el patrón de *no invarianza homogénea* a través de todas las condiciones, Δ_{RMSEA} muestra una mejora en la detección del patrón de *no invarianza homogénea*, a medida que aumenta el tamaño de la muestra y disminuye el número de variables, aunque en la mayoría de los casos los niveles de sensibilidad son insuficientes

Tabla 20.
Descriptivos para la diferencia en ajuste según Δ_{RMSEA} entre nivel configural y métrico para las distintas condiciones.

P	N	Invariante				No inv. homogéneo				No inv. mixto			
		Normativo		Ipsatizado		Normativo		Ipsatizado		Normativo		Ipsatizado	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
P=9	100	.000	.011	.001	.014	.020	.020	.006	.018	.028	.023	.013	.018
	300	.001	.006	.000	.008	.032	.011	.011	.013	.038	.012	.026	.014
	500	.000	.005	.000	.006	.035	.009	.016	.011	.042	.010	.030	.011
	1000	.000	.003	.000	.004	.037	.007	.020	.008	.044	.007	.034	.008
P=12	100	.000	.007	.000	.009	.013	.014	.007	.014	.028	.015	.021	.015
	300	.000	.005	.000	.006	.023	.009	.013	.009	.037	.009	.031	.010
	500	.000	.003	.000	.004	.026	.007	.017	.008	.040	.008	.034	.008
	1000	.000	.002	.000	.003	.028	.005	.020	.005	.042	.006	.036	.006
P=18	100	.001	.004	.000	.006	.009	.008	.005	.007	.022	.010	.020	.011
	300	.000	.003	.000	.003	.015	.006	.008	.006	.032	.007	.029	.007
	500	.000	.002	.000	.002	.017	.005	.009	.005	.034	.006	.031	.006
	1000	.000	.001	.000	.002	.019	.004	.012	.004	.036	.004	.033	.004

Nota. Los resultados están expresados en valor absoluto.

Tabla 21.
Proporción de soluciones en las cuales existe una pérdida significativa del ajuste según RMSEA al
ajustar el nivel métrico ($\Delta_{RMSEA} < .01$)

P	N	Invariante		No inv. homogéneo		No inv. Mixto	
		Norm.	Ips.	Norm.	Ips.	Norm.	Ips.
P=9	100	.102	.150	.553	.234	.778	.400
	300	.058	.083	.984	.470	.991	.871
	500	.040	.071	.994	.687	.997	.955
	1000	.008	.028	1.000	.913	1.000	1.000
P=12	100	.094	.112	.518	.310	.889	.710
	300	.046	.050	.951	.583	1.000	.990
	500	.006	.012	.998	.781	1.000	1.000
	1000	.000	.004	1.000	.961	1.000	1.000
P=18	100	.028	.064	.381	.235	.916	.836
	300	.002	.014	.788	.311	1.000	1.000
	500	.000	.000	.940	.443	1.000	1.000
	1000	.000	.000	.996	.669	1.000	1.000

Nota. Norm. = Normativo; Ips. = Ipsatizado.

En cuanto al índice *SRMR* los resultados difieren en relación a los obtenidos con *CFI* y *RMSEA*. Al comparar tanto la media como la proporción de rechazo del ajuste (véase tablas 22 y 23), se observa que Δ_{SRMR} alcanza valores inferiores en las condiciones ajustadas con datos *ipsatizados* que con datos *normativos*. Sin embargo, esta diferencia entre ambos tipos de datos va disminuyendo según aumenta el tamaño de la muestra. Esta pauta se repite en todas las condiciones de invarianza. Nótese además la excesiva tasa de error tipo II en ambos patrones de *no invarianza* (ver tabla 23). Estos resultados indican que el índice *SRMR* es demasiado conservador cuando se utiliza con datos *ipsatizados* para evaluar el nivel de invarianza métrico.

Tabla 22.
Descriptivos para la diferencia en ajuste según Δ_{SRMR} entre nivel configural y métrico para las distintas condiciones.

P	N	Invariante				No inv. homogéneo				No inv. mixto			
		Normativo		Ipsatizado		Normativo		Ipsatizado		Normativo		Ipsatizado	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
P=9	100	.028	.016	.017	.011	.069	.037	.026	.016	.044	.018	.022	.010
	300	.015	.007	.009	.005	.072	.017	.023	.010	.038	.009	.019	.007
	500	.011	.006	.007	.004	.075	.013	.025	.009	.041	.007	.019	.005
	1000	.008	.004	.005	.003	.079	.009	.027	.007	.043	.005	.020	.004
P=12	100	.024	.013	.014	.008	.065	.026	.026	.014	.038	.012	.025	.011
	300	.014	.007	.009	.005	.074	.017	.032	.012	.039	.007	.027	.008
	500	.011	.005	.007	.003	.079	.013	.035	.010	.041	.006	.028	.007
	1000	.007	.004	.004	.002	.085	.009	.039	.007	.044	.004	.032	.005
P=18	100	.023	.012	.012	.005	.070	.027	.020	.009	.038	.010	.020	.006
	300	.013	.006	.007	.003	.079	.016	.023	.007	.041	.006	.021	.004
	500	.010	.005	.005	.002	.082	.013	.024	.006	.043	.005	.023	.003
	1000	.007	.003	.004	.001	.089	.009	.028	.005	.048	.004	.027	.003

Nota. Los valores están expresados en valor absoluto.

En segundo lugar, en condiciones de no invarianza, los resultados de *SRMR* también contrastan con los de *CFI* y *RMSEA*. Mientras estos últimos dos índices mostraban dificultades para detectar la *no invarianza* con patrones *homogéneos*, *SRMR* presenta serias dificultades para detectar ambos tipos de no invarianza. Aunque existe un mayor número de condiciones con patrón mixto que alcanzan promedios por encima del punto de corte ($\Delta_{SRMR} > .025$), la proporción de rechazo del ajuste presenta niveles de potencia inaceptable tanto para el patrón de *no invarianza*

mixta como para el patrón de *no invarianza homogéneo*. Estos resultados muestran las importantes dificultades que posee el índice *SRMR* para detectar la no invarianza métrica frente a cualquiera de los patrones de invarianza.

Tabla 23
Proporción de soluciones en las cuales existe una pérdida significativa del ajuste según *SRMR* al ajustar el nivel métrico ($\Delta_{SRMR} < .01$)

<i>P</i>	<i>N</i>	<i>Invariante</i>		<i>No inv. homogéneo</i>		<i>No inv. Mixto</i>	
		<i>Norm.</i>	<i>Ips.</i>	<i>Norm.</i>	<i>Ips.</i>	<i>Norm.</i>	<i>Ips.</i>
<i>P=9</i>	<i>100</i>	.485	.208	.936	.468	.889	.356
	<i>300</i>	.105	.012	1.000	.402	.956	.208
	<i>500</i>	.020	.000	1.000	.468	.992	.115
	<i>1000</i>	.004	.000	1.000	.619	1.000	.127
<i>P=12</i>	<i>100</i>	.393	.088	.938	.473	.879	.430
	<i>300</i>	.069	.006	.998	.677	.980	.597
	<i>500</i>	.012	.002	1.000	.852	1.000	.635
	<i>1000</i>	.000	.000	1.000	.976	1.000	.922
<i>P=18</i>	<i>100</i>	.330	.022	.982	.261	.905	.185
	<i>300</i>	.050	.000	1.000	.349	.998	.192
	<i>500</i>	.012	.000	1.000	.437	1.000	.264
	<i>1000</i>	.000	.000	1.000	.729	1.000	.762

Nota. *Norm.* = Normativo; *Ips.* = Ipsatizado.

Para determinar el efecto de cada uno de los factores estudiados hemos realizado un ANOVA factorial para cada uno de los tres índices de ajuste. Para ello, y al igual que en los análisis anteriores, hemos utilizado la diferencia entre los valores obtenidos a nivel configural y el métrico. Los resultados se presentan en la tabla 24.

Tabla 24.

ANOVA factorial para la diferencia en los índices de ajuste Δ_{CFI} , Δ_{RMSEA} y Δ_{SRMR}

	gl	Δ_{CFI}			Δ_{RMSEA}			Δ_{SRMR}		
		F	p	η^2	F	p	η^2	F	p	η^2
D	1	23493.112	.000	.605	7632.539	.000	.333	79286.679	.000	.838
I	2	25845.235	.000	.796	23575.203	.000	.804	27533.345	.000	.782
N	3	93.477	.000	.018	716.296	.000	.123	69.212	.000	.013
P	2	146.659	.000	.019	498.032	.000	.061	106.194	.000	.014
D * I	2	7830.884	.000	.506	2678.563	.000	.259	22747.879	.000	.748
D * N	3	91.460	.000	.018	32.031	.000	.004	5.130	.002	.001
D * P	2	2643.512	.000	.257	649.713	.000	.078	894.980	.000	.105
I * N	6	51.522	.000	.020	244.264	.000	.087	460.379	.000	.156
I * P	4	575.713	.000	.131	218.683	.000	.054	61.479	.000	.016
N * P	6	35.898	.000	.014	13.323	.000	.005	23.227	.000	.009
D * I * N	6	25.660	.000	.010	20.313	.000	.008	116.511	.000	.044
D * I * P	4	1047.684	.000	.215	249.588	.000	.061	273.983	.000	.067
D * N * P	6	7.196	.000	.003	6.900	.000	.003	3.117	.005	.001
I * N * P	12	13.845	.000	.011	7.019	.000	.005	4.057	.000	.003

Nota. D = Tipo de datos; J = Nivel de invarianza; I = Tipo de invarianza; P = N° de variables y N = Tamaño de la muestra.

En cuanto a Δ_{CFI} , los efectos más importantes corresponden a los efectos principales de (I) patrón de invarianza ($\eta^2 = .796$) y de (D) tipo de datos ($\eta^2 = .605$), respectivamente. En cuanto a los efectos de interacción, los más relevantes corresponden a los efectos de doble interacción D*I ($\eta^2 = .506$), y D*P ($\eta^2 = .257$) y la triple interacción D*I*P ($\eta^2 = .215$). La figura 14 muestra en detalle los efectos de doble y triple interacción.

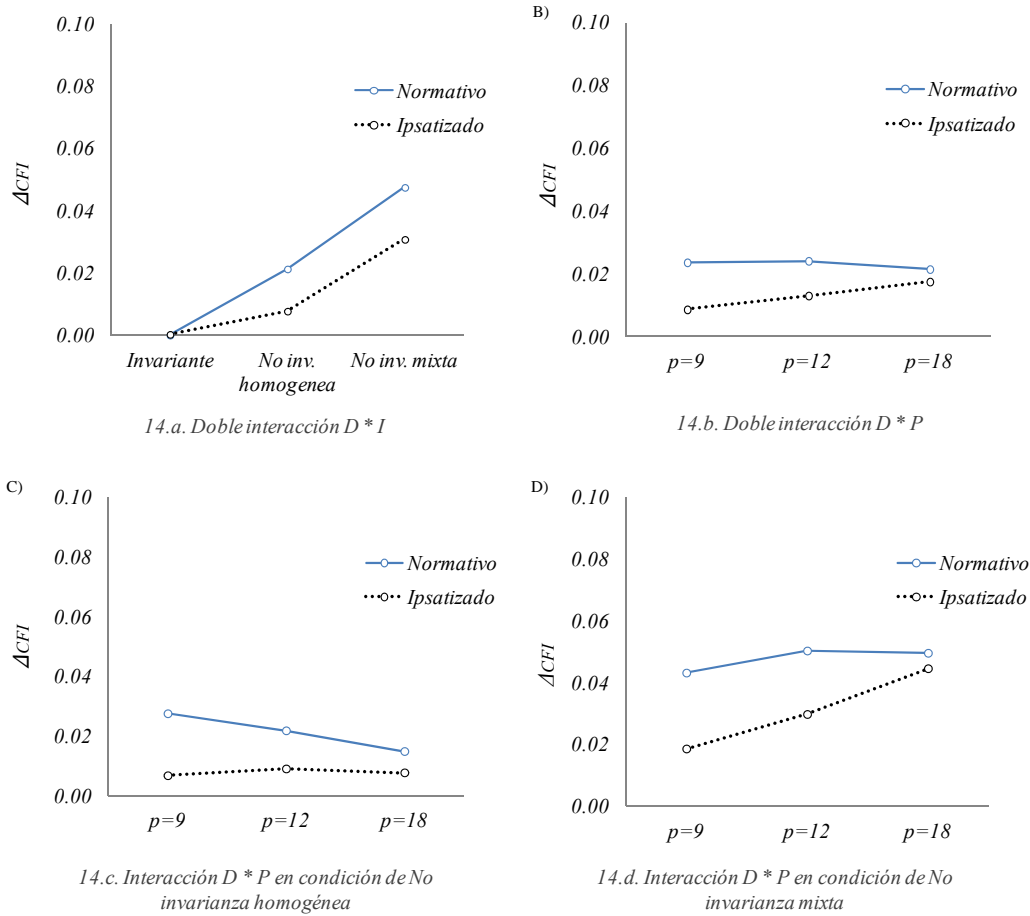


Figura 14. Efectos de interacción para Δ_{CFI}

Es posible observar que, tanto con datos *normativos* como con datos *ipsatizados*, Δ_{CFI} obtiene valores muy cercanos a cero en la condición de invarianza. Esto muestra que Δ_{CFI} logra detectar adecuadamente modelos invariantes entre los grupos, tanto con datos *normativos* como con datos *ipsatizados*. En cambio, cuando los modelos son no invariantes, Δ_{CFI} obtiene valores más altos cuando el patrón de invarianza es mixto que cuando es homogéneo, aunque en ambos casos estos valores son más bajos con datos *ipsatizados* que con datos *normativos*.

Por otro lado, independientemente de la condición de *invarianza*, la diferencia en Δ_{CFI} entre los datos *normativos* e *ipsatizados* es mayor cuando el modelo es pequeño ($P=9$), diferencia que disminuye significativamente a medida que aumenta el número de variables en el modelo (ver gráfico 14.b). Estos resultados muestran que, en condiciones de *invarianza*, el comportamiento de Δ_{CFI} es similar con ambos tipos de datos, detectando de manera adecuada la *invarianza* entre los grupos. Por otro lado, aunque con ambos tipos de datos Δ_{CFI} tiene dificultades para detectar la *no invarianza* cuando el patrón de *invarianza* es homogéneo, estas dificultades disminuyen a medida que aumenta el número de variables en el modelo (ver gráficos 14.c y 14.d).

En relación a Δ_{RMSEA} , el *patrón de invarianza* es el factor que posee un mayor efecto ($\eta^2 = .804$), seguido del *tipo de datos* ($\eta^2 = .333$), y de la interacción de ambos factores ($\eta^2 = .259$). Este último efecto se muestra en detalle en la figura 15. La gráfica muestra un fenómeno similar al que ocurre con *CFI*. Mientras que Δ_{RMSEA} obtiene valores muy cercanos a cero en condiciones de *invarianza* en ambos tipo de datos, Δ_{RMSEA} presenta dificultades para detectar la *no invarianza*, tanto más cuando el patrón es homogéneo que cuando es mixto. Nuevamente, existen diferencias entre los datos *normativos* e *ipsatizados* en ambas condiciones de *no invarianza*. A diferencia de *CFI*, el número de variables en el modelo no posee un efecto considerable en Δ_{RMSEA} . De forma similar a como ocurre con el índice *CFI*, el índice Δ_{RMSEA} detecta adecuadamente a los grupos invariantes, ya sea con datos *normativos* o con datos *ipsatizados*. Sin embargo, presenta dificultades para detectar

grupos *no invariantes*, las cuales se ven incrementadas con datos *ipsatizados* y el patrón de invarianza sea *homogéneo*. A diferencia de *CFI*, el funcionamiento de *RMSEA* no mejora con el aumento del número de variables en el modelo.

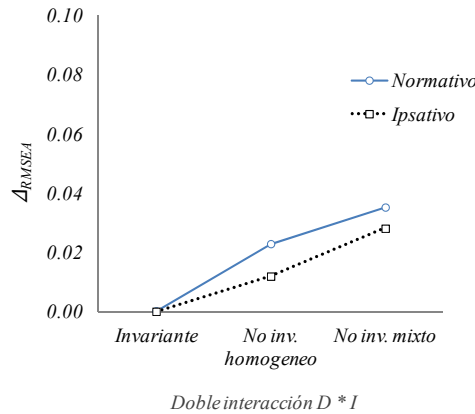


Figura 15. Efectos de doble interacción para Δ_{RMSEA}

Al igual que en el análisis descriptivo, el comportamiento de Δ_{SRMR} difiere de los anteriores índices de ajuste. A pesar de verse afectado por los mismos efectos principales de *tipo de datos* ($\eta^2 = .838$) y el *patrón de invarianza* ($\eta^2 = .782$), el modo en que interaccionan es considerablemente distinto. A diferencia de Δ_{CFI} y Δ_{RMSEA} , que no difieren cuando los grupos son *invariantes*, Δ_{SRMR} para datos *normativos* obtiene valores mayores en todos los patrones de invarianza. Este ocurre incluso cuando los grupos son *invariantes*, aunque tanto datos *normativos* como *ipsatizados* alcanzan medias muy por debajo del criterio que permite rechazar el ajuste ($\Delta_{SRMR} = .025$). Por otro lado, en ambas condiciones de invarianza, los datos *normativos* alcanzan valores más altos en Δ_{SRMR} que los datos *ipsatizados*, aunque a

diferencia de *CFI* y *RMSEA*, las mayores diferencias en *SRMR* aparecen cuando el patrón de invarianza es homogéneo. Por otro lado, las mayores diferencias entre los tipos de datos se encuentran precisamente frente a la condición de no invarianza homogénea.

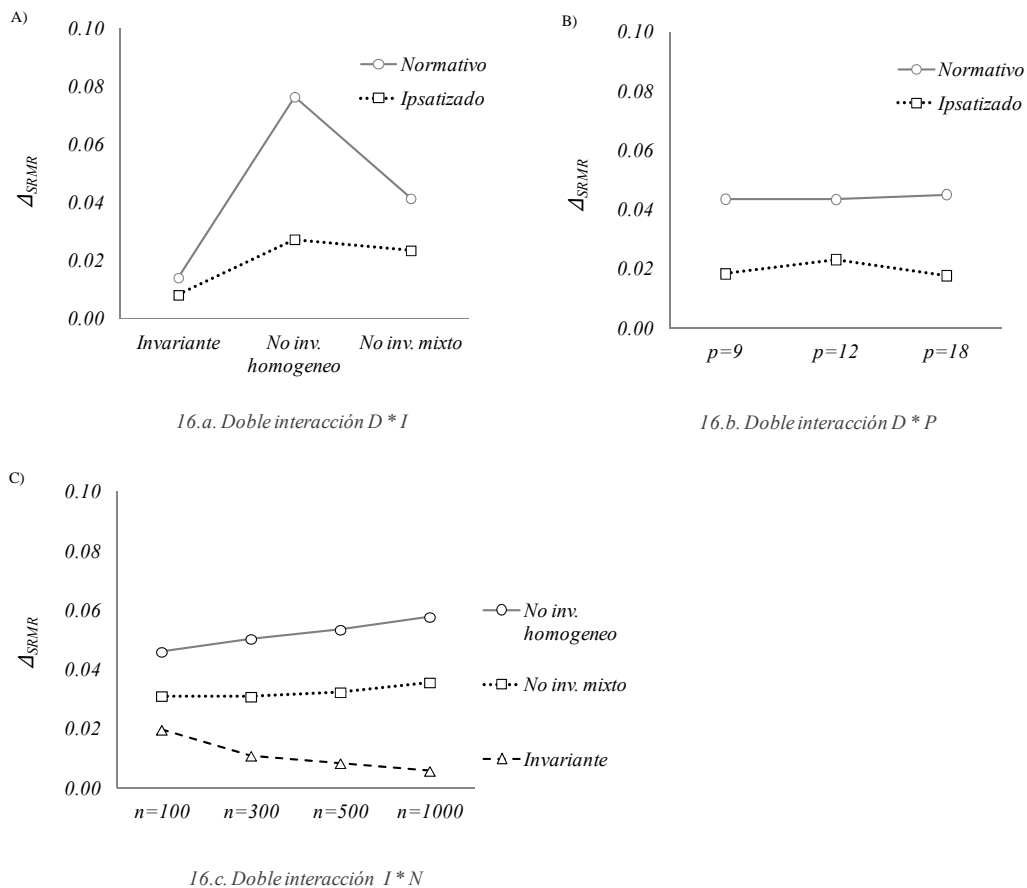


Figura 16. Efectos de doble interacción para *SRMR*

Finalmente, otros efectos de interacción considerables corresponden a las interacciones dobles $I*N$ ($\eta^2 = .156$) y $D*P$ ($\eta^2 = .105$), aunque estos efectos son más pequeños. De ellos, el que parece más destacable corresponde a la interacción

*I*N*. La gráfica 16.c muestra en detalle este efecto. Podemos ver que, independientemente del *tipo de datos*, Δ_{SRMR} se ve afectado por el tamaño de la muestra de forma distinta, dependiendo del patrón de invarianza. Cuando los grupos son invariantes los valores Δ_{SRMR} disminuyen en la medida que aumenta el tamaño de la muestra. Por otro lado, cuando los grupos son no invariantes, Δ_{SRMR} aumenta a medida que aumenta el tamaño de la muestra, siendo mayor cuando el patrón de invarianza es homogéneo. Esto ocurre en todos los tamaños muestrales.

5.2.5. Conclusiones

Los resultados del estudio 2 son similares a los del estudio 1 y a los hallazgos previos con datos *ipsatizados*. En presencia de grupos *invariantes*, los índices de ajuste logran detectar de manera adecuada la invarianza métrica (Cheung, 2004; Chan y Bentler, 1993; 1996;), mientras que cuando los grupos son no invariantes, estos índices se tornan conservadores, presentando dificultades para detectar las diferencias en las saturaciones factoriales. Este resultado es coherente con lo hallado por Ximénez y Calderón (2012). En algunos casos, y de manera similar a los hallazgos previos con datos *normativos* (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004), estos problemas dependerán del patrón de no invarianza presente en las diferencias en las saturaciones factoriales entre los grupos.

En el caso del estadístico χ^2 , el contraste de la hipótesis de invarianza métrica obtiene tasas de error tipo I dentro de los valores aceptables, detectando adecuadamente la invarianza métrica entre los grupos. No obstante, cuando los

grupos son no invariantes, la potencia del estadístico χ^2 depende del patrón de invarianza. Cuando existe un patrón de invarianza homogéneo, χ^2 se torna excesivamente conservador teniendo problemas para detectar la no invarianza métrica, especialmente cuanto menor sea el tamaño de la muestra y menor el número de variables en el modelo. En cambio, cuando el patrón de no invarianza es mixto, χ^2 sólo presenta problemas con modelos con pocas variables ($P=9$) y tamaños muestrales pequeños ($N=100$), detectando adecuadamente la no invarianza métrica en el resto de las condiciones. Estos resultados son similares a los obtenidos en la investigación previa con datos *normativos* (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004), en donde el patrón de invarianza fue uno de los factores que mejor explica la variabilidad de los índices de ajuste.

En relación al índice *CFI* ocurre una pauta similar. Aunque detecta adecuadamente la invarianza métrica con datos *ipsatizados*, se torna excesivamente conservador cuando el patrón de no invarianza es homogéneo, no logrando detectar este patrón de no invarianza en ninguna de las condiciones de P o N . Por otro lado, y al igual que χ^2 , logra detectar adecuadamente la no *invarianza mixta*, a excepción de los modelos con pocas variables ($P=9$) y tamaños de la muestra reducidos ($N=100$), aunque este último no logra un tamaño del efecto importante.

Con *RMSEA*, los resultados son similares a los obtenidos con *CFI* y χ^2 , con algunas diferencias. Aunque en el ANOVA no se mostró un efecto importante del tipo de datos sobre Δ_{RMSEA} , la proporción de rechazo de ajuste nos muestra que la

potencia mejora cuando el número de variables es reducido y el tamaño de la muestra es mayor.

Finalmente, el índice *SRMR* presenta los resultados más dispares en comparación con el resto de los índices de ajuste. Los resultados muestran que al comparar los resultados entre los dos tipos de datos, *SRMR* se torna excesivamente conservador con datos *ipsatizados*, independientemente de si los grupos son invariantes o no invariantes, a pesar del efecto grande del patrón de invarianza (*I*). Aunque las interacciones *D*P* y *N*I* presentan efectos considerables, estos no se ven reflejados en una mejora en la tasa de error tipo II. Esto nos lleva a la conclusión de que, a diferencia de los hallazgos previos reportados con datos *normativos* (Chen, 2007), *SRMR* no logra detectar la *no invarianza* métrica con datos *ipsatizados* independientemente del patrón de no invarianza existente entre los grupos.

Estos resultados pueden tener explicación desde la investigación previa (Chen, 2007). Mientras el índice *CFI* y *RMSEA* han mostrado mayor consistencia a través de los distintos niveles de invarianza y mayor sensibilidad para detectar la falta de invarianza de las saturaciones factoriales, el índice *SRMR* posee un funcionamiento diferencial, mostrando ser más sensible a las variaciones muestrales cuando se utiliza en el análisis a nivel métrico. Es posible que la pérdida de sensibilidad del índice *SRMR* a nivel métrico con variables *ipsatizadas* se vea empeorada por la pérdida de información que sufre la matriz de covarianzas durante el proceso de factorización de variables *ipsatizadas*.

A diferencia de la investigación previa con variables *normativas* (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004), el empeoramiento de

los índices de bondad de ajuste es mucho mayor en algunas condiciones para *CFI* y *RMSEA*, concretamente en modelo con pocas variables ($P=9$) y en tamaños muestrales reducidos ($N=100$), y en especial para el índice *SRMR*, el cual alcanza niveles de sensibilidad inaceptables en todas las condiciones, independientemente del patrón de no invarianza presente en las saturaciones factoriales entre los grupos.

5.3. SÍNTESIS GENERAL Y RECOMENDACIONES PRÁCTICAS DE LOS ESTUDIOS DE SIMULACIÓN

El propósito de los estudios de simulación presentados fue el de evaluar el funcionamiento de los índices de bondad de ajuste más empleados en la práctica para el estudio de la invarianza de medida con datos *ipsatizados*; particularmente en lo referido a la sensibilidad para detectar la falta de invarianza factorial, así como determinar las condiciones óptimas para llevar a cabo este tipo análisis.

Nos hemos centrado especialmente en los dos niveles de invarianza de mayor interés en la práctica investigadora: el nivel invarianza configural o modelo base, y el nivel invarianza métrico, debido, fundamentalmente, a que el cumplimiento de la invarianza en estos dos niveles suele considerarse condición suficiente para mostrar la equivalencia de medida, en el análisis comparativo de distintos grupos.

Los resultados proporcionan información sugerente en la comprensión del funcionamiento de los índices de bondad de ajuste con variables *ipsatizadas* en el análisis factorial en general, y en el análisis factorial multi-grupo en particular. Por otro lado, están dentro de los resultados esperados y son coherentes con lo hallado en

investigaciones previas con datos *ipsatizados* y con el AFCM con datos *normativos*. Estos resultados cobran relevancia debido a que son los primeros estudios que se llevan a cabo en la evaluación del AFCM con este tipo de datos.

Una de las conclusiones más importantes que se deriva de ambos estudios tiene relación con la pérdida de sensibilidad de los índices de bondad de ajuste para detectar errores de especificación en el AFC de variables *ipsatizadas*. Mientras que en condiciones de invarianza entre los grupos, el funcionamiento de los índices de bondad de ajuste es similar al de los datos *normativos*, la sensibilidad de los índices de ajuste se ve seriamente afectada con datos *ipsatizados* en condiciones de no invarianza. Este fenómeno ocurre tanto a nivel configural, como a nivel métrico. Por tanto, es recomendable que los investigadores procedan con cautela al llevar a cabo un AFCM con datos *ipsatizados*, especialmente si se sospecha de una falta de invarianza entre los grupos, sea ésta a nivel configural o métrico.

A continuación, y de acuerdo a los resultados obtenidos, presentamos una serie de recomendaciones prácticas para llevar a cabo de manera adecuada un análisis de equivalencia de medida a través de un AFCM con variables *ipsatizadas*. Siguiendo el esquema del análisis de equivalencia/invarianza de la medida, las recomendaciones van dirigidas en primer lugar al ajuste del modelo configural o base, y posteriormente las referidas al nivel métrico. Debido a que los resultados encontrados indican que en condiciones de invarianza entre los grupos, el funcionamiento de los índices de ajuste no presenta diferencias importantes entre datos *normativas* e *ipsatizados*, las recomendaciones están dirigidas especialmente cuando se tienen sospechas de la existencia de falta de invarianza entre los grupos.

Los resultados muestran que el funcionamiento de los índices de bondad de ajuste con datos *ipsatizados* presentan problemas importantes, tanto de confianza como de sensibilidad, cuando las muestras son pequeñas (e.g., $N < 100$). Por este motivo, con *datos ipsatizados* se recomienda el uso de muestras de tamaño algo mayor (e.g., $N \geq 300$) al requerido con datos *normativos*. Esta condición es necesaria tanto para el nivel de invarianza configural como para el métrico. En segundo lugar, debido a que en el proceso de factorización de variables *ipsatizadas* es necesaria la eliminación de una de las variables para evitar la singularidad de la matriz de covarianzas, no es recomendable utilizar modelos con un número reducido de variables (e.g., $P = 9$) ya que la pérdida de información es proporcionalmente mayor en modelos con pocas variables. Por tanto, se recomienda emplear modelos con un número mayor de variables (e.g., $P \geq 12$).

Una vez recogidos los datos, el primer paso es evaluar el ajuste del modelo configural o base. Para detectar adecuadamente la falta de invarianza configural es necesario seguir una serie de recomendaciones. En primer lugar, en una muestra de 300 o más observaciones, el estadístico χ^2 muestra un comportamiento similar que el que presenta con datos *normativos*, por lo que el contraste de la hipótesis de igualdad entre los grupos se debe llevar a cabo de manera similar que con datos *normativos*.

En segundo lugar, los estadísticos *CFI* y *SRMR* sólo deben utilizarse cuando el número de variables en el modelo sea mayor que 9, ya que los modelos con pocas variables presentan elevadas tasas de error tipo II. En modelos con pocas variables es factible la utilización del índice *RMSEA*, aunque su interpretación se debe llevar a cabo con cautela ya que el número de variables no ha mostrado un efecto importante.

Aunque *RMSEA* muestra resultados algo mejores en muestras pequeñas, en modelos con muchas variables debe ser utilizado con cautela, ya que ha mostrado problemas para detectar grupos no invariantes (e.g., en modelos con 18 variables).

Finalmente, los resultados muestran que al evaluar el ajuste del nivel de invarianza métrico el comportamiento de los índices de ajuste con datos *ipsatizados* mejora substancialmente, llegando a ser muy similar que con datos *normativos*. Cabe destacar, como veremos más adelante, que el uso convencional de los índices de bondad de ajuste obtenidos a nivel métrico con datos *ipsatizados* sólo aseguraría la invarianza configural. Para determinar la *invarianza métrica* con datos *ipsatizados* es necesario otro tipo consideraciones, las cuales veremos a continuación.

Una vez asumida la *invarianza configural*, el siguiente paso corresponde al ajuste del nivel de invarianza más restrictivo: el nivel de *invarianza métrico*. Los resultados muestran que, en general, el funcionamiento de los índices de ajuste con datos *ipsatizados* en este nivel, es similar al observado a nivel configural. Mientras que con ambos tipos de datos los índices de ajuste logran detectar de manera adecuada los grupos que son invariantes, estos muestran dificultades para detectar la no invarianza de las saturaciones factoriales cuando el modelo se ajusta con datos *ipsatizados*. Sin embargo, estos problemas de sensibilidad dependerán del tipo de patrón de no invarianza que presenten los grupos.

Cabe señalar que, a diferencia de lo que sucede con datos *normativos*, frente a los cuales el estadístico *SRMR* muestra mayor sensibilidad a nivel métrico, los resultados obtenidos con datos *ipsatizados* muestran elevadas tasa de error tipo II, independientemente del patrón de invarianza presente en los grupos. Por tanto, se

desaconseja el uso de este índice para la evaluación del ajuste a nivel métrico con datos *ipsatizados*, ya que este estadístico se torna excesivamente conservador frente a cualquier patrón de no invarianza.

Cuando el patrón de no invarianza es mixto la mayoría de los índices de ajuste, a excepción de *SRMR*, poseen niveles de sensibilidad adecuados. Al igual que para el nivel configural, tanto el estadístico χ^2 como *CFI*, poseen problemas para detectar la no invarianza en modelos con pocas variables y muestras pequeñas (por ejemplo, $P=9$ y $N=100$), por lo que en el resto de las condiciones su interpretación procede de igual manera que con datos normativos. Cabe destacar que para el caso de *CFI* la sensibilidad mejora considerablemente a medida que aumenta el número de variables en el modelo.

A diferencia que con el nivel configural, a nivel métrico el índice *RMSEA* no presenta pérdida de sensibilidad al aumentar el número de variables en el modelo, por lo que es igualmente recomendable en presencia de un mayor número de variables. Los resultados del ANOVA muestran que el índice *RMSEA* no se ve afectado por el número de variables, por lo que se puede utilizar cualquiera que sea el tamaño del modelo, aunque al igual que *CFI*, no se recomienda su uso con modelos con pocas variables ($P=9$) y con un número reducido de observaciones ($N=100$).

Por el contrario, cuando el patrón de no invarianza es homogéneo, todos los índices de ajuste empeoran sensiblemente su funcionamiento. Mientras el índice *CFI* posee graves problemas de sensibilidad en todas las condiciones, los índices *SRMR* y *RMSEA* obtienen niveles de sensibilidad adecuada sólo con tamaños muestrales excesivamente grandes ($N=1000$), algo muy difícil de cumplir en la práctica

investigadora. Por tal motivo, se recomienda que los investigadores analicen detalladamente las estimaciones de las soluciones obtenidas. Si se detecta un patrón de invarianza homogéneo, se sugiere abandonar el análisis y descartar el modelo, independientemente de los valores de los índices de ajuste obtenidos.

CAPITULO VI

UN ESTUDIO EMPIRICO:

Comparación de la bondad de ajuste de dos formatos de presentación del Inventario de Estilos de Aprendizaje, LSI de Kolb

En este último capítulo, hemos considerado un estudio con datos reales con un doble propósito. El primero de ellos es ilustrar el procedimiento del análisis de invarianza factorial con datos *ipsatizados*, utilizando de método *DE* de Cheung (2004) y su implementación en un programa SEM. Y en segundo lugar, comparar el funcionamiento del análisis de invarianza factorial para variables *ipsatizadas* que fueron obtenidas en un formato *normativo* tradicional y su posterior *ipsatización*, con los resultados obtenidos de un conjunto de datos que fueron recogidos de forma *ipsativa*, a través del formato de *respuesta forzada*.

6.1. MÉTODO

6.1.1. *Instrumento*

El instrumento correspondió a una versión reducida del “Inventario de Estilos de Aprendizaje”, LSI (*Learning Style Inventory*) de Kolb (1985) que se muestra en el anexo 6. La versión original está compuesta de 12 bloques de preguntas ($p=12$), de 4 alternativas en cada bloque ($n=4$), lo que hace un total de 48 ítems en total ($m=48$). Cada una de las alternativas del bloque evalúa una de las cuatro estrategias de aprendizaje, que en combinación definen los cuatro estilos de aprendizaje definidos por el test: *acomodador*, *divergente*, *asimilador* y *convergente*.

En nuestro estudio hemos considerado la utilización de una versión reducida de sólo 4 bloques. Esta decisión obedeció fundamentalmente a dos criterios: en primer lugar, los test con formato de *respuesta forzada*, al ser codificados como comparaciones binarias, requieren de la estimación de la matriz asintótica para la estimación del modelo. Es por ello que la utilización del test completo hubiese requerido de una muestra excesivamente grande para la adecuada estimación de la matriz asintótica. En segundo lugar, con la presentación de los 48 ítems que conforman el test en los dos formatos, se corría el riesgo de producir agotamiento en los entrevistados, provocando una pérdida de fiabilidad en las medidas.

Para la elección de los bloques hemos realizado un estudio piloto, intentando seleccionar los ítems con mejores propiedades psicométricas. El análisis está basado fundamentalmente en los resultados hallados en un análisis factorial exploratorio previo sobre la versión *normativa* del test.

El diseño del estudio consideró la presentación de la versión reducida del test en dos formatos de respuesta. A cada participante le fue presentado en primer lugar el test en su formato original, el cual correspondió al formato de respuesta forzada, constituido por cuatro bloques ($p=4$) de 4 alternativas cada uno ($n=4$), conformando un total de 16 ítems. Al entrevistado se le dio la instrucción de ordenar cada una de las 4 alternativas de cada bloque desde el 4 (“la alternativa que más te representa”), hasta el 1 (“la alternativa que menos te representa”). A continuación, una vez completado el test en el formato de *respuesta forzada*, se administró el test en el formato *normativo*. Los mismos 16 ítems que conformaron los cuatro bloques del formato de *respuesta forzada* fueron presentados cada uno con una escala Likert de 5 puntos (de *totalmente en desacuerdo* a *totalmente de acuerdo*). Ambos conjuntos de

datos fueron analizados con el propósito de evaluar el funcionamiento del análisis de invarianza factorial sobre variables *ipsatizadas*.

6.1.2. Muestra

Un total de 593 estudiantes de la Universidad Católica del Norte (Antofagasta, Chile), respondieron ambos formatos de presentación del test. La mayoría de participantes correspondieron a estudiantes de primer año de psicología (52.6%). El resto pertenecían a primer año de titulaciones de ingeniería (28.4%), arquitectura (10.1%), periodismo (5.9%) y economía (3.0%). El rango de edad fue de 17 y 38 años, con una media de 19.80 y una desviación típica de 2.36 años. En cuanto al sexo, 260 estudiantes fueron hombres (43.8%) y 333 fueron mujeres (56.2%). La variable sexo se utilizó como variable de agrupación en el análisis de la invarianza factorial.

6.1.3. Procedimiento

Una vez obtenidos los dos conjuntos de datos a partir de los dos formatos de respuesta, hemos realizado tres análisis. En primer lugar hemos realizado un AFCM convencional sobre los datos *normativos* obtenidos con el formato de respuesta graduada (Likert). Posteriormente los datos *normativos* fueron transformados a su versión *ipsatizada* a través de la *transformación ipsativa* descrita en la ecuación [1.10]. Los datos normativos fueron analizados con un AFC convencional y los datos *ipsatizados* con el método *DE* de Cheung (2003). En ambos casos hemos utilizado el método de Máxima Verosimilitud con el programa EQS v6 (Bentler, 2004). Para la identificación del modelo hemos fijado la varianza de los factores latentes a 1, como

sugieren Chan y Bentler (1993) y Cheung (2004). Para el caso de los datos *ipsatizados*, hemos fijado la varianza de los factores de 2° orden, los cuales corresponden a los factores de primer orden del modelo *normativo* convencional. El anexo 2 presenta la sintaxis en EQS utilizada. Finalmente, hemos llevado a cabo un AFCM con los datos *ipsativos* producto del formato de *respuesta forzada*, utilizando el modelo TRI Thurstoniano de Brown y Maydeu-Olivares (2011). El modelo factorial correspondió a una estructura de cuatro factores con cuatro ítems por factor, los cuales fueron presentados en cuatro bloques. Cada uno de los ítems del bloque está orientado a medir uno de los cuatro factores. Los cuatro ítems de cada bloque fueron codificados a través de 6 comparaciones binarias, las cuales se vinculan a cada una de las cuatro *utilidades* latentes correspondientes a cada ítem (de manera similar a la estructura presentada en la figura 5). El anexo 2 presenta la sintaxis utilizada².

Para estimar el modelo hemos utilizado el método WLSM incorporado en el programa Mplus v5 (Múthen y Múthen, 2006; ver también nuestro anexo 3).

Para la identificación global del modelo hemos fijado la varianza de los factores y una de los errores de las *utilidades* latentes en cada bloque a 1. Del mismo modo que los estudios de simulación, para los tres tipos de datos hemos ajustado sólo los modelos a nivel de *invarianza configural* (base) y a *nivel de invarianza métrico* (anidado). Para la conformación de los distintos grupos hemos utilizado la variable sexo como variable de agrupación.

² El lector puede consultar la web de la autora: <http://annabrown.name/software>, para descargar la macro Exel y su correspondiente manual de usuario para crear y editar sintaxis de Mplus para modelos TRI Thurstoniano de test con formato de respuesta forzada, especificando el número y diseño de los bloques.

6.2. RESULTADOS

La tabla 25 muestra los índices de bondad de ajuste para los tres conjuntos de datos a nivel configural, así como la diferencia en el ajuste con el nivel de invarianza métrico.

Tabla 25.
Índices de bondad de ajuste para los tres conjuntos de datos a nivel de invarianza configural y a nivel de invarianza métrica

	Formato de respuesta graduada (Likert)						Formato de respuesta forzada		
	Datos normativos			Datos ipsatizados			Datos ipsativos		
	CF	MC	Δ	CF	MC	Δ	CF	MC	Δ
χ^2	431.2	442.6	11.5	312.3	326.9	14.6	2002.9 ¹	2007.4 ¹	4.87
<i>Gl</i>	196	212	16	164	180	16	512 ²	528 ²	16
<i>p-valor</i>	.000	.000	.799	.000	.000	.553	.000	.000	.996
<i>CFI</i>	.906	.908	.002	.950	.950	.000	.922	.922	.000
<i>RMSEA</i>	.057	.063	.006	.058	.055	.002	.098	.100	.002
<i>SRMR</i>	.068	.070	.002	.064	.067	.003	----	----	----

Nota. CF=Nivel de invarianza configural; MC=Nivel de invarianza métrica; Δ =diferencia en el ajuste entre el nivel configural y el métrico.

¹El estadístico de bondad de ajuste está ajustado con la corrección Satorra-Bentler (1994).

²Los grados de libertad están corregido debido a las redundancias existentes en la matriz asintótica.

En primer lugar, los resultados en los índices de bondad de ajuste a nivel configural del modelo ajustado con datos *normativos* muestran que existe un ajuste moderado. Aunque el estadístico χ^2 (431.2; $p<.000$) nos lleva al rechazo de la hipótesis de ajuste, el resto de los índices de ajuste no muestran evidencia unívoca en relación a la falta de invarianza de los grupos. A pesar de que *CFI* muestra un ajuste

moderado (valores entre .90-.95), los índices *SRMR* ($<.08$) y *RMSEA* ($<.60$) muestran un ajuste aceptable. Al analizar la diferencia en el ajuste con el modelo anidado (invarianza métrica), ninguno de los índices muestra una pérdida significativa del ajuste al imponer las restricciones de igualdad en las saturaciones factoriales.

Tal y como recomiendan Cheung y Chan (2002), bajo la sospecha de la existencia de un *sesgo de respuesta* que afecta a los resultados del análisis de la invarianza, el siguiente paso es comparar estos resultados del ajuste con los obtenidos con el AFCM sobre los datos *ipsatizados*. Podemos ver que al ajustar el modelo con datos *ipsatizados* el valor de índice *CFI* mejora considerablemente, llegando a un valor que informa de un buen ajuste del modelo (.95). El resto de los índices de ajuste no muestran cambios considerables en comparación con el modelo *normativo*. Tanto el índice *RMSEA* como el índice *SRMR* se mantienen dentro de valores que indican la existencia de invarianza entre los grupos, en especial este último, el cual recordemos, en el estudio de simulación no mostró pérdida de sensibilidad en condiciones de no invarianza configural.

Estos resultados sugieren la existencia de un efecto de un *sesgo de respuesta* sobre los resultados del modelo *normativo*, el cual puede estar siendo eliminado a través de la *transformación ipsativa*. Como hemos sugerido a partir de los estudios de simulación, es necesario interpretar estos resultados con cautela, debido a la pérdida de sensibilidad de los índices de ajuste en el AFCM con datos *ipsatizados*.

Al observar los valores obtenidos a *nivel métrico*, puede verse que ninguno de los índices muestra una pérdida significativa en el ajuste. Esto contrasta con los hallazgos encontrados en el estudio de simulación en la condición de no invarianza, el

cual mostró que los valores de los índices de bondad de ajuste con datos *ipsatizados* cambian significativamente entre el nivel de invarianza configural y el nivel de invarianza métrica, poniendo en evidencia la falta de invarianza. Estos resultados confirman la existencia de *invarianza configural* entre los grupos.

Aunque no son concluyentes, estos resultados muestran la existencia de un ajuste relativo de los datos a la invarianza del modelo entre los grupos. Por otro lado, pueden ser evidencia del efecto de un *sesgo de respuesta* presente en los datos *normativos*, el cual se ve en parte eliminado tras la *ipsatización*. Finalmente, recordemos que, aunque nos hemos valido de los resultados del ajuste de *invarianza métrica*, no es posible de momento, concluir la existencia de invarianza a este nivel. Para ello es necesario realizar una exploración detallada de las estimaciones de las saturaciones factoriales para determinar el patrón de invarianza existente.

La tabla 26 muestra los parámetros estimados sobre los datos *normativos* en el modelo de invarianza configural en ambos grupos. Si observamos en detalle el valor de los parámetros, podemos advertir la clara existencia de un patrón mixto de diferencias en las saturaciones factoriales entre hombres y mujeres. Para el caso del grupo de hombres, los coeficientes con valores más altos corresponden a los ítems 1, 4, 8, 9, 10, 11, 13 y 16, y para el grupo de mujeres los ítems 2, 3, 7, 12 y 15.

Una vez determinado el patrón de diferencias en las saturaciones factoriales es posible interpretar los índices de bondad de ajuste en el análisis de invarianza de los grupos a nivel métrico. Puede verse que tanto con datos *normativos* como con datos *ipsatizados* no existe una pérdida significativa del ajuste entre el nivel de *invarianza configural* y el nivel de *invarianza métrica* ($p=.799$ y $p=.553$ respectivamente). En

cuanto al resto de los índices de bondad de ajuste, tanto los índices Δ_{CFI} ($<.01$), Δ_{RMSEA} ($<.01$) como Δ_{SRMR} ($<.025$), muestran que no existe pérdida significativa de ajuste al imponer la restricción de igualdad en las saturaciones factoriales entre los grupos. Esto ocurre con ambos tipos de datos.

Tabla 26.
Parámetros estimados del modelo normativos a nivel configural en ambos grupos

	Hombres			Mujeres				Hombres			Mujeres		
	λ	St.	S.E.	λ	St.	S.E.		λ	St.	S.E.	λ	St.	S.E.
y_1	.840	.766	.074	.782	.751	.060	y_9	.762	.849	.055	.724	.801	.053
y_2	.710	.662	.072	.734	.702	.060	y_{10}	.491	.565	.057	.389	.468	.051
y_3	.682	.662	.070	.755	.702	.062	y_{11}	.649	.751	.054	.574	.645	.053
y_4	.432	.432	.071	.402	.438	.057	y_{12}	.547	.548	.066	.593	.632	.056
y_5	.499	.636	.049	.499	.648	.042	y_{13}	.340	.374	.062	.319	.403	.048
y_6	.608	.759	.048	.619	.731	.045	y_{14}	.636	.819	.054	.633	.826	.044
y_7	.533	.712	.046	.575	.745	.041	y_{15}	.327	.348	.064	.433	.469	.055
y_8	.780	.838	.054	.674	.819	.042	y_{16}	.643	.865	.052	.583	.819	.041

Nota. λ =parámetro no estandarizado; St=parámetro estandarizado; S.E.=error de estimación

Aunque los tres índices de ajuste presentan resultados concordantes, es importante considerar que el estadístico de ajuste Δ_{RMSEA} debe interpretarse con cautela, ya que como hemos visto, este estadístico sufre una pérdida considerable de la sensibilidad al contrastar la igualdad de las saturaciones factoriales con datos *ipsatizados*. Pese a ello, el resto de los resultados indican la existencia de invarianza en las saturaciones factoriales entre ambos grupos, tanto con datos *normativos* como con datos *ipsatizados*.

Estos resultados indican que, bajo el supuesto de la existencia de un *sesgo de respuesta* sobre los datos normativos en el AFCM, la *ipsatización* del conjunto de datos logra eliminar el impacto del *sesgo de respuesta* en el análisis de la E/IM entre los grupos.

Con el propósito de buscar mayor evidencia acerca de la efectividad de la *ipsatización* de variables en el tratamiento del impacto del *sesgo de respuesta* en el AFCM, hemos comparado los resultados del AFCM de datos *ipsatizados* con los resultados obtenidos del ajuste de la estructura factorial con los datos del test presentado en formato de *respuesta forzada*.

Los resultados muestran que el ajuste de los datos de *respuesta forzada* es pobre. Tanto el estadístico χ^2 ($p < .000$) como el índice de ajuste *RMSEA* ($> .06$) llevan al rechazo de la hipótesis de ajuste. Sólo el índice *CFI* muestra evidencia de un ajuste moderado, aunque no sobrepasa el criterio para ser un ajuste aceptable ($CFI \geq .95$). Pese a ello, los tres estadísticos de ajuste muestran que no existe una pérdida significativa de ajuste al restringir las saturaciones factoriales, aunque como se ha visto, no es posible asumir la invarianza a nivel configural.

Por tanto, estos resultados muestran que el modelo TRI Thurstoniano no ajusta a los datos obtenidos a partir del test presentado en formato de *respuesta forzada*. Este resultado puede deberse fundamentalmente a dos causas principales. La primera de ellas puede ser el error en la especificación del modelo. El ajuste del modelo AFC con datos *normativos*, como el de datos *ipsatizados*, sugieren que la falta de ajuste del test de *respuesta forzada* no se debe al error de especificación, sino a que el modelo TRI Thurstoniano y el sistema de codificación binaria, no logran

recuperar de manera satisfactoria la estructura subyacente. En segundo lugar, debido a que el modelo TRI Thurstoniano requiere la estimación de la matriz asintótica, el método de estimación exige requisitos demasiado restrictivos (por ejemplo, tamaños muestrales más grandes que los modelos con variables continuas), o condiciones de difícil cumplimiento en la práctica (por ejemplo, ítems con dirección opuesta). Debido a ello, es posible que los resultados obtenidos se deban al incumplimiento de estos requisitos y/o condiciones, y no al error en la estructura subyacente a los datos.

En cualquier caso, creemos que a la luz de los resultados obtenidos en este caso, la *ipsatización* de variables es una mejor alternativa que el formato de *respuesta forzada* para el tratamiento del *sesgo de respuesta*, al menos hasta que no se encuentren hallazgos que garanticen el mejor funcionamiento de los test de *respuesta forzada* y el modelo TRI Thurstoniano.

6.3. CONCLUSIONES

El presente estudio ha tenido un doble propósito. Por un lado, ilustrar el procedimiento de ajuste de un modelo AFCM con datos *ipsatizados* y su implementación en un programa SEM. Por otro, estudiar la efectividad de la *ipsatización* de variables para eliminar el impacto del *sesgo de respuesta* en un conjunto de datos. Para ello, hemos utilizado datos empíricos obtenidos de una versión reducida del “Inventario de Estilos de Aprendizaje” de Kolb (1985) presentada en dos formatos de respuesta. La primera de ellas correspondió a la forma de presentación original del test: el formato de *respuesta forzada*. La segunda fue una versión del mismo test, esta vez en formato *normativo* (Likert). Hemos ajustado tres

modelos (*normativo*, *ipsatizado* e *ipsativo de respuesta forzada*), cuyos resultados fueron comparados para evaluar la efectividad del AFCM con variables *ipsatizadas* a través del método *DE* de Cheung (2004).

Los resultados obtenidos nos entregan información sugerente acerca de la implementación del AFCM de variables *ipsatizadas* y de la efectividad de la *ipsatización* de variables en el tratamiento del impacto del *sesgo de respuesta* en el análisis de E/IM. Siguiendo las recomendaciones de Cheung y Chan (2002), hemos comparado los índices de bondad de ajuste del modelo *normativo*, asumiendo la ausencia de sesgo, y del modelo *DE* con datos *ipsatizados*, asumiendo la presencia de sesgo.

Los resultados encontrados pueden ser evidencia de la efectividad de la *ipsatización* para tratar el *sesgo de respuesta*. En primer lugar, al comparar los resultados obtenidos a nivel configural entre datos *normativos* y datos *ipsatizados*, podemos ver que el modelo de datos *ipsatizados* posee mejor ajuste que el modelo ajustado con datos *normativos*. Además, al ajustar el nivel de invarianza métrica no existe una pérdida significativa del ajuste, lo cual apoya la existencia de invarianza a nivel configural.

En segundo lugar, la existencia de un patrón mixto de diferencias en las saturaciones factoriales permite continuar con el análisis a nivel métrico. Además, la ausencia de una pérdida significativa del ajuste a este nivel, apoyan la existencia de E/IM del formato *normativo* entre hombres y mujeres.

Finalmente, la falta de ajuste del modelo a partir de los datos de *respuesta forzada* muestra que el modelo TRI Thurstoniano falla al recuperar la estructura

interna de los datos. Aunque es posible que la falta de ajuste del modelo no se deba al error en la estructura subyacente hipotetizada, sino más bien a las restricciones del método de estimación (por ejemplo, debido al tamaño de la muestra), o del modelo propiamente dicho (por ejemplo, ítems con la misma dirección) que afectan a la adecuada estimación del modelo.

En síntesis, nuestros resultados sugieren que, sea cual fuese la causa de la falta de ajuste, la *ipsatización* de variables y el posterior análisis a través del método *DE* ofrecen los medios más adecuados para analizar la E/IM a través del AFCM cuando se asume la existencia de un impacto del *sesgo de respuesta* en los datos. Especialmente, cuando no se cumplen algunas de las condiciones requeridas por el modelo TRI Thurstoniano (por ejemplo, las características de los bloques), del método de estimación y, particularmente, cuando las características del test dificultan la implementación de condiciones deseables para el ajuste y estimación adecuados de modelos de TRI Thurstoniano en test de *respuesta forzada*, como puede ser un número reducido de dimensiones o ítems.

CAPITULO VII

CONCLUSIONES Y COMENTARIOS FINALES

El propósito de esta tesis ha sido estudiar la efectividad de la *ipsatización* de variables para el tratamiento del impacto del *sesgo de respuesta* y el funcionamiento de algunos de los índices de ajuste más utilizados en el AFCM en el contexto del análisis de la equivalencia/invarianza de medida (E/IM) entre distintos grupos. El trabajo ha sido motivado fundamentalmente por dos razones. En primer lugar, debido a la capacidad potencial de la *ipsatización* para eliminar el efecto del *sesgo de respuesta* en el estudio de la invarianza factorial de medida y la falta de estudios que evalúen su efectividad. Y en segundo lugar, por la escasez de estudios, hasta el momento, que intentan evaluar las condiciones óptimas para el uso del AFCM con variables *ipsatizadas*.

La primera parte del trabajo se ha destinado a presentar los elementos teóricos fundamentales del uso de la *ipsatización* para el tratamiento del *sesgo de respuesta* en el contexto de la E/IM y el fundamento analítico del AFC de variables *ipsatizadas*, junto con su extensión al análisis multi-grupo. También se ha presentado una de las estrategias más utilizadas desde el punto de vista instrumental para el tratamiento del *sesgo de respuesta*: los test de formato de *respuesta forzada*. Hemos utilizado este formato de test en el estudio empírico con el propósito de evaluar el funcionamiento y efectividad de la *ipsatización* para el tratamiento del *sesgo de respuesta*. Finalmente,

hemos presentado una revisión de los escasos estudios existentes que han evaluado algunas condiciones óptimas para llevar a cabo un AFC con variables *ipsatizadas*. Hasta la fecha, son pocos los estudios que examinan las condiciones óptimas para la aplicación de la factorización de datos *ipsatizados* y que ofrecen recomendaciones prácticas a los investigadores sobre cómo deben diseñar sus estudios para poder llevar a cabo la factorización de datos *ipsatizados* con ciertas garantías, y ninguno de estos estudios ha considerado el modelo AFCM. Uno de los problemas asociados a la escasez de trabajos publicados donde se implemente el AFC de variables *ipsatizadas* puede deberse a la complejidad del planteamiento teórico y la dificultad en su implementación, incluso a pesar de que la técnica de análisis se haya simplificado bastante desde que fue inicialmente formulada (Chan y Bentler, 1993), hasta la actual propuesta (Cheung, 2004).

En los capítulos introductorios hemos intentado sintetizar los trabajos de Chan (2003), Chan y Bentler (1993; 1996) y Cheung (2003; 2004), para el caso de AFC de datos *ipsatizados* y hemos puesto a disposición del lector las sintaxis de los programas SEM utilizados para el ajuste de esta clase de modelos. Además, hemos presentado el procedimiento para llevar a cabo un análisis factorial de datos producto de un test de formato de *respuesta forzada* a través del modelo TRI Thurstoniano (Brown y Maydeu-Olivares, 2011). Para más detalles sobre estos procedimientos el lector puede remitirse a las referencias originales³.

³ Para una revisión en detalle del método *DE* para variables *ipsatizadas*, el lector puede consultar a Cheung (2004). Para detalles sobre la implementación del método TRI Thurstoniano, remitirse a Brown y Maydeu-Olivares (2011) y a la web <http://annabrown.name/software>.

Además de servir para el análisis del problema y el planteamiento de los objetivos e hipótesis de investigación, esta primera parte de la tesis tenía como propósito ilustrar al lector en los procedimientos para llevar a cabo un AFCM de datos con propiedades *ipsativas*, tanto si han sido obtenidos en formato *normativo* y posteriormente *ipsatizados*, como si han sido obtenidos directamente de forma *ipsativa* (formato de *respuesta forzada*).

En la segunda parte de la tesis se han presentado los objetivos del trabajo, las hipótesis y un plan de trabajo para estudiar los objetivos propuestos. El propósito principal del presente trabajo ha sido el de estudiar la efectividad de la *ipsatización* de variables como estrategia para evitar el impacto del *sesgo en la respuesta* en el estudio de la invarianza factorial mediante modelos de AFC multi-grupo o modelos AFCM. Para ello, hemos llevado a cabo dos estudios de simulación y un estudio empírico, los cuales tenían como objetivos examinar el funcionamiento y la efectividad del AFCM con datos *ipsatizados* en el tratamiento del *sesgo de respuesta*.

En relación a los estudios de simulación, los resultados han sido coherentes con la hipótesis 1 referida al funcionamiento de los índices de bondad de ajuste en el AFC de variables *ipsatizadas*, así como al funcionamiento general del AFCM con variables *normativas*. Globalmente, y al igual que en los estudios anteriores (Chan, 2004; Chan y Bentler, 1993, 1996), frente a grupos invariantes, los índices de bondad de ajuste en el AFCM de variables *ipsatizadas* poseen un comportamiento similar que con variables *normativas*. Por el contrario, cuando los grupos son no invariantes, o lo

que es lo mismo, frente a errores de especificación, el funcionamiento general de los índices de bondad de ajuste empeora, perdiendo sensibilidad para detectar la falta de invarianza entre los grupos. Esto es coherente con los resultados hallados en el estudio de Ximénez y Calderón (2012).

Dentro de los factores estudiados, los que mostraron un mayor efecto sobre el funcionamiento de los índices de bondad de ajuste con datos *ipsatizados* fueron el nivel de invarianza, el número de ítems en el modelo y el patrón de las diferencias en las saturaciones factoriales. Uno de los factores no analizados en estudios anteriores y que mostró un efecto estadísticamente significativo en el funcionamiento del AFCM de datos *ipsatizados* es el nivel de invarianza. Los resultados muestran que el funcionamiento de los índices de bondad de ajuste empeora al ajustar un modelo con datos *ipsatizados* en comparación con el ajuste obtenido con datos *normativos*. Esta diferencia entre datos *normativos* e *ipsatizados* disminuye considerablemente cuando se imponen las restricciones de igualdad de las saturaciones factoriales.

Tal como fue planteado en la hipótesis 2, y de manera similar a los que ocurre con el AFCM de datos *normativos* (Chen, 2007; Meade y Bauer, 2007; Meade y Lautenschlager, 2004), el patrón de las diferencias factoriales es uno de los factores que mayor impacto tiene sobre la sensibilidad de los índices de bondad de ajuste a nivel métrico. Como habíamos hipotetizado, la pérdida de sensibilidad para detectar la falta de invarianza es mucho mayor cuando el patrón de las diferencias en las saturaciones factoriales es homogéneo que cuando es mixto.

Por otro lado, como muestra la investigación previa (Ximénez y Calderón, 2012) y tal como fue planteado en la hipótesis 3, el número de variables en el modelo posee un efecto estadísticamente significativo en la sensibilidad de los índices de bondad de ajuste para detectar el error de especificación. Más concretamente, la pérdida de sensibilidad de los índices de bondad ajuste es mayor cuanto menor es el número de ítems en el modelo. Ello se justifica en que al aumentar el número de variables en el modelo, la pérdida de información, tras la eliminación de una variable de la matriz de covarianzas, es proporcionalmente menor.

Finalmente, en cuanto al estudio empírico, siguiendo las recomendaciones de Cheung y Chan (2002), hemos comparado los resultados en los índices de bondad de ajuste tras ajustar el modelo de AFCM convencional con datos *normativos* y el modelo de AFCM con el método *DE* de Cheung (2004) para los datos *ipsatizados*. Los resultados indican que la mejora en los índices de bondad de ajuste del modelo con datos *ipsatizados*, en comparación con los datos *normativos*, sugiere la presencia de un impacto del *sesgo de respuesta* sobre los resultados obtenidos con datos *normativos*, y que la *ipsatización* logra eliminar. Junto con ello, el ajuste relativo del modelo TRI Thurstoniano de datos *ipsativos de respuesta forzada* da apoyo también a la efectividad de la *ipsatización* en la eliminación del impacto del *sesgo de respuesta* en el estudio de la E/IM. Sin perjuicio de lo anterior, se necesita mayor investigación que apoye los resultados con respecto al funcionamiento del AFCM de variables *ipsatizadas* y la efectividad en el tratamiento del *sesgo de respuesta* en estudios de E/IM.

Implicaciones de los resultados e investigación futura

Los resultados obtenidos permiten ofrecer una serie de recomendaciones prácticas para los investigadores.

En primer lugar, se recomienda a los investigadores que planteen *ipsatizar* sus datos bajo la sospecha de la presencia de *sesgo de respuesta*, que tengan ciertas precauciones a la hora de diseñar sus estudios, particularmente en relación a la determinación del número de variables en el modelo y el tamaño de la muestra. Dado que en el AFC con variables *ipsatizadas* en general, es necesario eliminar una variable de la matriz de covarianzas de entrada para evitar su singularidad, se pierde información durante el proceso. Por tanto, se recomienda considerar un mayor número de variables para que la pérdida de información sea proporcionalmente menor. La inclusión de un mayor número de variables en el modelo no sólo influye en la potencia y sensibilidad de los índices de bondad de ajuste, sino también, como demostraron Ximénez y Calderón (2012), mejora la recuperación de las saturaciones factoriales y disminuye los errores de estimación. Por otro lado, se recomienda utilizar tamaños muestrales algo más grandes que los requeridos en el ajuste de modelos con variables *normativas* (por ejemplo, $N \geq 300$), ya que, como hemos visto, en tamaños muestrales reducidos, los índices de bondad de ajuste pierden sensibilidad para detectar la falta de invarianza.

En segundo lugar, los resultados hallados en el presente trabajo sugieren que, una vez obtenidos los datos, es necesario determinar la presencia de *sesgo de respuesta*. Siguiendo las recomendaciones de Cheung y Chan (2002), se debe comparar el ajuste del modelo con datos *normativos*, ignorando la presencia de sesgo,

con el del modelo AFCM con datos *ipsatizados*, asumiendo la presencia de sesgo. Una mejora en el ajuste con datos *ipsatizados*, respecto de los datos *normativos*, es evidencia de la presencia de un efecto del *sesgo de respuesta* sobre los resultados obtenidos con los datos *normativos*. Tras determinar la mejora en el ajuste del modelo con datos *ipsatizados* y asumir la presencia de *sesgo de respuesta*, es posible continuar con el análisis invarianza con los datos *ipsatizados*.

A partir de los resultados de los estudios de simulación es posible realizar una serie de recomendaciones para llevar a cabo un AFCM con datos *ipsatizados* con ciertas garantías.

1. Al ajustar el nivel de invarianza configural o base, tanto el estadístico χ^2 como el índice *CFI*, pueden utilizarse e interpretarse de la misma forma que con datos *normativos*, a excepción de modelos que posean un número reducido de variables (e.g., $P = 9$) y con tamaños de la muestra muy pequeños (e.g., $N = 100$). Por el contrario, cuando el modelo posee pocas variables (e.g., $P = 9$), el índice *RMSEA* posee un mejor funcionamiento, aunque para modelos con mayor número de variables (e.g., $P = 18$) su interpretación requiere de cierta cautela. En cuanto al índice *SRMR* sólo se recomienda su uso con modelos de doce o más variables ($P \geq 12$), ya que sufre una pérdida significativa de sensibilidad en modelos con pocas variables, independientemente del tamaño muestral.
2. Para apoyar la existencia de invarianza configural entre los grupos, se recomienda la imposición de la restricción de igualdad de las saturaciones factoriales. Los resultados muestran que, en presencia de no invarianza a nivel

configural entre los grupos, los índices de ajuste con datos *ipsatizados* poseen un comportamiento similar a los de los datos *normativos*. Es importante considerar que el ajuste del nivel de invarianza métrico directamente, no asegura la existencia de invarianza de las saturaciones factoriales. Para ello es necesario previamente examinar el patrón de diferencias en las saturaciones. Es decir, para apoyar el adecuado ajuste de la invarianza a nivel configural, se necesita revisar los resultados del ajuste del invarianza a nivel métrico, si llegar a concluir la existencia de invarianza a este nivel.

3. Una vez asumida la existencia de invarianza a nivel configural, el primer paso en el análisis de la invarianza a nivel métrico es el examen de las saturaciones factoriales para determinar el patrón de diferencias existentes. La presencia de un patrón de diferencias homogéneo entre las saturaciones factoriales de los grupos conlleva una pérdida significativa de la sensibilidad para todos los índices de bondad ajuste. Por tanto, frente a un patrón de *no invarianza homogéneo*, es recomendable desechar inmediatamente el modelo, ya que los índices de bondad de ajuste no garantizan la detección de la falta de invarianza entre los grupos. En cambio, si se detecta un patrón de invarianza mixto, es posible continuar con la evaluación de la invarianza métrica entre los grupos de manera similar que a nivel configural. Tanto la diferencia en el ajuste según χ^2 , Δ_{CFI} y Δ_{RMSEA} , pueden utilizarse en la mayoría de las condiciones, a excepción de modelos con pocas variables (e.g., $P=9$) y con muestras pequeñas (e.g., $N=100$). A diferencia del nivel de invarianza configural, al imponer las restricciones de igualdad en el nivel de invarianza métrica, no es recomendable el uso de índice de ajuste Δ_{SRMR}

ya que se torna excesivamente conservador, independientemente del patrón de la diferencias en las saturaciones factoriales.

En cuanto a la efectividad de la *ipsatización* de variables para el tratamiento del *sesgo de respuesta*, los resultados del estudio empírico ofrecen algunas sugerencias.

1. En primer lugar, la mejora en los índices de bondad de ajuste al ajustar el modelo con datos *ipsatizados*, en comparación de los datos *normativos*, sugiere la presencia de un impacto del *sesgo de respuesta* que afecta a la recuperación del modelo con datos *normativos*, y que la *ipsatización* logra eliminar.
2. En segundo lugar, la falta de ajuste de los datos obtenidos con el test con formato de *respuesta forzada*, muestran que el modelo TRI Thurstoniano no logra recuperar la estructura interna de los datos. Los buenos resultados obtenidos tanto por el modelo *normativo* como por el *ipsatizado* permiten asumir que la estructura subyacente hipotetizada es la estructura correcta, y que la falta de ajuste del modelo TRI Thurstoniano se debe a su incapacidad de recuperar tal estructura. Es posible que esta incapacidad se deba al incumplimiento de algunas de las condiciones necesarias para el adecuado funcionamiento del modelo TRI Thurstoniano. Algunas de estas condiciones (por ejemplo, la redacción de ítems con dirección opuesta dentro de cada bloque) son de difícil implementación en la práctica y pueden dificultar el proceso de respuesta de los entrevistados. Es por ello que creemos necesario llevar a cabo más investigación que intente resolver los problemas en el ajuste de esta clase de modelos. Hasta que no sean

desarrolladas mejoras o nuevas propuestas que incorporen condiciones menos restrictivas o más realistas, la *ipsatización* de variables se convierte en la estrategia más eficiente y de más fácil aplicación en situaciones reales para el tratamiento del *sesgo de respuesta* en los estudios de E/IM.

3. Las evidencias acerca de la efectividad de la *ipsatización* de variables para eliminar el efecto del *sesgo de respuesta* proporcionadas por nuestros estudios no son concluyentes, y se hacen necesarios más estudios para confirmar estos hallazgos. Estos esfuerzos deben estar orientados a la implementación de diseños que incorporen la posibilidad de controlar y/o modelar el efecto del *sesgo de respuesta*. Un ejemplo de esta clase de diseños es el implementado en Stark et al (2001), en el cual se estudió el efecto del falseamiento de respuesta sobre el funcionamiento de un test de personalidad. El diseño consideró el control del falseamiento de respuesta por medio de la manipulación de las situaciones de evaluación de los distintos grupos evaluados. Un diseño experimental de estas características permitiría determinar y controlar el impacto del *sesgo de respuesta*, y junto con ello, aportar mayor evidencia sobre la efectividad de la *ipsatización* para su tratamiento.
4. Por otro lado, los investigadores que se planteen la comparación de las propiedades psicométricas de datos *normativos* e *ipsatizados* de un mismo test, deben ser conscientes de que los índices de bondad de ajuste no son directamente comparables (Chan y Bentler, 1993). Esto ocurre especialmente por el cambio en los grados de libertad entre el modelo *normativo* y el *ipsatizado*. Ello hace que no exista una relación directa entre modelos con ambos tipos de datos.

5. Otra dirección en el estudio del AFC de variables *ipsatizadas* se refiere a la evaluación de otros factores que han mostrado un efecto importante en el AFCM con variables normativas. Como en todo estudio, las conclusiones extraídas a partir de los resultados obtenidos no pueden generalizarse a otro tipo de modelos o condiciones no consideradas en nuestro diseño. Ejemplo de condiciones o factores que han sido estudiadas en trabajos que evalúen el funcionamiento del AFCM y que podrían ser extendidos al AFCM con variables *ipsatizadas* son: la proporción de no invarianza a nivel métrico (cantidad de ítems que presentan diferencias en las saturaciones entre grupos), la magnitud de las saturaciones factoriales, magnitud de las comunalidades, fiabilidad de los ítems, entre otras.
6. Mención aparte merece el problema del análisis de la invarianza parcial (Byrne, 1989). Dado que la asunción de igualdad de la totalidad de un conjunto de parámetros (por ejemplo, la totalidad de la matriz de saturaciones factoriales) es un requisito muy exigente en la práctica, algunos investigadores consideran razonable liberar algunas de las restricciones de igualdad entre-grupos en alguno de los parámetros. Un problema potencial que trae consigo la asunción de la invarianza parcial a nivel métrico, está relacionado con la necesidad de fijar la métrica de los factores latentes. Sea utilizando un ítem de referencia o fijando la varianza de los factores, ambas estrategias son demasiado arriesgadas, ya que bajo el supuesto de invarianza parcial no se podría asumir la igualdad del parámetro de referencia o la igualdad de las varianzas de los factores latentes.
7. En cuanto al modelo de AFC para variables *ipsatizadas*, aún son muchas las cuestiones que se necesitan resolver para su adecuada implementación. Un problema fundamental es que las puntuaciones *ipsativas* no son comparables

inter-individualmente, lo que hace que su uso sea limitado. Aunque Cheung (2003) propone la estimación de las puntuaciones factoriales, es necesaria más evidencia para verificar su utilidad.

8. Finalmente, un problema aún no resuelto es el de la identificación del modelo *ipsatizado*. Como advierten Chan y Bentler (1993, 1996), los investigadores que utilicen la *ipsatización* en sus estudios deben ser conscientes de que un modelo *pre-ipsativo (normativo)* es condición necesaria, pero no suficiente para la identificación de su correspondiente modelo *ipsatizado*. Por ejemplo, en un AFC convencional un modelo de un factor puede identificarse, sin embargo su correspondiente modelo *ipsatizado* no es identificable (Jackson y Alwin, 1980). Es por ello que los investigadores que se propongan trabajar con datos *ipsatizados* deben ser cautelosos en el examen de la identificación local de un modelo particular.

En síntesis, y a pesar de que no son concluyentes, los resultados obtenidos en este trabajo apoyan la efectividad de la *ipsatización* en el tratamiento del *sesgo de respuesta* y sugieren recomendaciones prácticas para la adecuada implementación del AFCM de variables *ipsatizadas* en el contexto de E/IM. Por tanto, dado el potencial de esta técnica, creemos importante continuar con la investigación sobre el funcionamiento del AFCM de variables *ipsatizadas*.

REFERENCIAS

- Aitchison, J. (1986) The statistical analysis of compositional data. New York: Chapman y Hall. En Bekker, P. (1989). Identification in restricted factor models and the evaluation of rank conditions. *Journal of econometrics*, 41, 5-16.
- Alwin, D. F. y Jackson, D. J. (1981). Applications of simultaneous factor analysis to issues of factorial invariance. In D. Jackson y E. Borgatta (Eds.), *Factor analysis and measurement in sociological research: A multi-dimensional perspective*. Beverly Hills, CA: Sage.
- Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, 69, 49-56.
- Bartram, D. (2007). Increasing Validity with Forced- Choice Criterion Measurement Formats. *International Journal of Psychological Assessment*, 15, 263-272.
- Bentler, P. M. (1990). Comparative fit indices in structural equation models. *Psychological Bulletin*, 107, 238–246.
- Bentler, P. M. (2004). *EQS 6: Structural equation program manual*. Encino, CA: Multivariate Software.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care*, 44, 176–181.
- Brown, A y Maydeu-Olivares, A. (2011). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, 71, 460-504.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A. y Kaemmer, B. (1989). *MMPI-2: Manual of administration and scoring*. Minneapolis, MN: University of Minnesota Press.
- Byrne, B. M. (1988). The Self-Description Questionnaire III: Testing for equivalent factorial validity across ability. *Educational and Psychological Measurement*, 48, 397–406.

- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling*, 11, 272-300.
- Byrne, B. M. y Cambell, T. L. (1999). Cross-Cultural Comparisons and the Presumption of Equivalent Measurement and Theoretical Structure: A Look Beneath the Surface. *Journal of Cross-Cultural Psychology*, 30, 555-574.
- Byrne, B. M. y Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of second-order factor structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, 13, 287-321.
- Byrne, B. M., Shavelson, R. J., y Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456-466.
- Cattell, R. (1944). Psychological measurement: Ipsative, normative and interactive. *Psychological Review*, 51, 291-303.
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating longitudinal mean and covariance structures analysis (LMACS) and multiple indicator latent growth modeling (MLGM). *Organizational Research Methods*, 1, 421-483.
- Chan, D. y Schmitt (1997). Video-based versus paper and pencil method of assessment in situational judgment test: subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 85, 190-210
- Chan, D. y Schmitt, N. (2000). Interindividual differences in intraindividual changes in proactivity during organizational entry: A latent growth modeling approach to understanding newcomer adaptation. *Journal of Applied Psychology*, 85, 190-210.

- Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, 30, 99-121.
- Chan, W. y Bentler, P. M. (1993). The covariance structure analysis of ipsative data. *Sociological Method y Research*, 22, 214-247.
- Chan, W. y Bentler, P. M. (1996). Covariance structure analysis of partially ipsative data using restricted maximum likelihood estimation. *Multivariate Behavioral Research*, 31, 289-312.
- Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- Cheung, G. W. y Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*. 25, 1-27.
- Cheung, G. W. y Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Cheung, M. (2004). A direct estimation method on analyzing ipsative data with Chan and Bentler's (1993) method. *Structural Equation Modeling*, 11, 217-243.
- Cheung, M. (2006). Recovering preipsative information from additive ipsatized Data: A Factor Score Approach. *Educational and Psychological Measurement*, 66, 565-588.
- Cheung, M. y Chan, W. (2002): Reducing uniform response bias with ipsative measurement in multiple-group confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 55-77
- Chun, K. T., Campbell, J. B., y Yoo, J. H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, 5, 465-480.
- Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, 69, 41-47.

- Cornwell, J. M. y Dunlap, W. P. (1994). On the questionable soundness of factoring of ipsative data: A response to Seville y Willson (1991). *Journal of Occupational and Organizational Psychology*, 67, 89-100.
- Cunningham, W. H., Cunningham, I. M., y Green, R. T. (1977). The Ipsative Process to Reduce Response Set Bias. *Public Opinion Quarterly*, 41, 379-388.
- Dimitrov, D. (2010). Testing for factorial invariance in the context of construct validity. *Measurement and Evaluation in Counseling and Development* 43, 121-149.
- Drasgow, F. (1987), Study of measurement bias of two standardized psychological test. *Journal of Applied Psychology*, 72, 19-29.
- Drasgow, F., y Kanfer, R. (1985). Equivalent of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70, 662-680.
- Dunlap, W. P. y Cornwell, J. M. (1994). Factor analysis of ipsative measures. *Multivariate Behavioral Research*. 29,115-126.
- Facteau, J. D., y Craig, S. B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86, 215-227.
- Fava, J. L. y Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement*, 56, 907-929.
- Fisher, R. (2004). Standardization to Account for Cross-Cultural Response Bias: A Classification of Score Adjustment Procedures and Review of Research in JCCP. *Journal of Cross-Cultural Psychology*, 35, 263-282.
- French, B. y Finch, W. (2006). Confirmatory Factor Analytic Procedures for the Determination of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 13, 378-402.
- Golembiewski, R. T., Billingsley, K. y Yeager, S. (1975). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133-157.

- Greenleaf, E. A. (1992a). Improving rating scale measures by detecting and correcting bias components in some response styles. *Journal of Marketing Research*, 29, 176–188.
- Grimm, S. D., y Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, 33, 415–441.
- Haladyna, T. M., y Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23, 17–27.
- Hicks, L. E. (1970). Some properties of ipsative, normative and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- Horn, J. L. y McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144
- Hu, L-T. y Bentler, P. (1998). Fit indices in the covariance factor structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 4, 179-188.
- Hu, L-T. y Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation modeling*, 6, 1-55.
- Jackson, D. J., y Alwin, D. F. (1980). The factor analysis of ipsative measures. *Sociological Methods y Research*, 9, 218–238.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.
- Jöreskog, K. G., y Sörbom, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: ABT Books.

- Jöreskog, K. G., y Sörbom, D. (1989). *LISREL 7: User's reference guide*. Chicago: Scientific Software.
- Kaplan, D. (1989). The problem of error rate inflation in covariance structure models. *Educational and Psychological Measurement*, 49, 333–337.
- Kolb, D. A. (1985). *The learning style inventory: technical manual*. Boston: McBer.
- Labouvie, E. y Ruetsch, C. (1995). Testing for equivalence of measurement scales: Simple structure and metric invariance reconsidered. *Multivariate Behavioral Research*, 30, 63–76.
- León, O. y Montero, I. (2001). Cómo explicar el concepto de interacción sin estadística: análisis gráfico de todos los casos posibles en un diseño 2 x 2. *Psicothema*, 13, 159-165.
- León, O. y Montero, I. (2003). *Métodos de investigación en psicología y educación*. Madrid: McGraw-Hill/Interamericana de España S.A.
- Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53–76.
- Loo, R. (1999). Issues in factor-analyzing ipsative measures. The learning style inventory (LSI-1985) example. *Journal of business and Psychology*, 14, 149-151.
- Marquis, K. H., Marquis, M. S. y Polich, J. M. (1986). Response bias and reliability in sensitive topic surveys. *Journal of the American Statistical Association*, 81, 381–389.
- Marsh, H. W. (1985). The structure of masculinity/femininity: An application of confirmatory factor analysis to higher-order factor structures and factorial invariance. *Multivariate Behavioral Research*, 20, 427–449.
- Maydeu-Olivares, A., y Böckenholt, U. (2005). Structural equation modeling of paired comparisons and ranking data. *Psychological Methods*, 10, 285-304.

- Meade, A. W. y Bauer, D. J. (2007). Power and precision in confirmatory factor analysis test of measurement invariance. *Structural Equation Modeling*, 14, 611-635.
- Meade, A. W. y Kroustalis, C. M. (2006). Problem with item parceling for confirmatory factor analytic test of measurement invariance. *Organizational Research Methods*, 9, 369-403.
- Meade, A. W. y Lautenschlager, G. L. (2004). A Monte-Carlo study of confirmatory factor analytic test of measurement equivalence/invariance. *Structural Equation Modeling*, 11, 60-72.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13, 127-143.
- Meredith, W. (1993). Measurement invariance, factor analysis and factor invariance. *Psychometrika*, 58, 525-543.
- Meredith, W. y Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57, 289-311.
- Meredith, W. y Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, 69-77.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed). New York, NY: American Council on Education.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Millsap, R. E. (1995). Measurement invariance, predictive invariance, and the duality paradox. *Multivariate Behavioral Research*, 30, 577-605.
- Millsap, R. y Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*. 9, 93-115.
- Múthen, L.K. y Múthen, B.O. (2006). *MPLUS. Statistical analysis with latent variables: User's guide*. Los Angeles, CA: Muthén and Muthén.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598–609.
- Paulhus, D. L. (1991). Measures of personality and social psychological attitudes. In J. P. Robinson y R. P. Shaver (Eds.), *Measures of social psychological attitudes series* (Vol. 1, pp. 17–59). San Diego: Academic.
- Ployhart, R. E., Weekley, J. A., Holtz, B. C. y Kemp, C. (2003). Web-based and paper-and-pencil testing of applicants in a proctored setting: Are personality, biodata and situational judgment tests comparable? *Personnel Psychology*, 56, 733–752.
- Raju, N. Laffitte, L. y Byrne, B. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87, 517–529.
- Riordan, C. M. y Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management*, 20, 643–671.
- Schmitt, N. (1982). The use of analysis of covariance structure to assess the beta and gamma change. *Multivariate Behavioral Research*, 17, 343–358.
- Schmitt, N. y Kuljanin. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18, 210–222.
- Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: attacking the conventional wisdom. *Multivariate Behavioral Research*, 35, 137–167.
- Stark, S., Chernyshenko, O. S., Chan, K. Y., Lee, W. C., y Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, 86, 943–953.
- Steiger, J. H. y Lind, J. C. (1980). Statistically based tests for the number of common factors. Trabajo presentado en el *Annual meeting of the Psychometric Society*, Iowa City, IA.

- Taris, T. W., Bok, I. A. y Meijer, Z. Y. (1998). Assessing stability and change of psychometric properties of multi-item concepts across different situations: A general approach. *Journal of Psychology*, 132, 301-316.
- Thurstone, L. L. (1931). Rank order as a psychological method. *Journal of Experimental Psychology*, 14, 187-201.
- Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Technical Report 984). Washington, DC: Department of the Army.
- Van de Vijver, F. y Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
- Vandenberg, R. J. y Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-69.
- Vandenberg, R. J. y Self, R. M. (1993). Assessing newcomers' changing commitment to the organization during the first 6 months of work. *Journal of Applied Psychology*, 78, 557-568.
- Ximénez, C. (2006). A Monte Carlo Study of Recovery of weak factor loadings in Confirmatory Factor Analysis. *Structural Equation Modeling*, 13, 587-614.
- Ximénez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under model misspecification conditions. *Behavior Research Methods*, 41, 1038-1052.
- Ximénez, C. y Calderón, C. (2012). El análisis factorial de datos ipsativos: un estudio de simulación. *Psicothema*, 24, 302-309.
- Yoon, M. y Millsap, R. (2007). Detecting violations of factorial invariance using data-based specification searches: a Monte Carlo study. *Structural Equation Modeling*, 14, 435-463.

ANEXOS

TERMINOLOGÍA

AFC	Análisis Factorial Confirmatorio
AFCM	Análisis Factorial Confirmatorio Multi-grupo
E/IM	Equivalencia/Invarianza de Medida
y	Vector de variables normativas observadas
y^*	Vector de variables normativas con sesgo de respuesta
τ	Vector de interceptos
ξ	Vector de factores latentes
ε	Vector de errores de medida
κ	Vector de medias de los factores latentes
Σ	Matriz de covarianzas poblacional
S	Matriz de covarianzas observada
Λ	Matriz de saturaciones factoriales de 1º orden
Φ	Matriz de covarianzas de los factores latentes
Ψ	Matriz de covarianzas de los errores de medida del modelo normativo

ANEXO 1

Sintaxis EQS para la estimación del modelo de 3 factores de 3 ítems por factor

```

/TITLE
Sintaxis EQS para análisis de invarianza con datos ipsatizados entre dos grupo utilizando el método
DE, 3 factores, 3 variables por factor-grupo 1;
/SPECIFICATIONS
DATA= «*.dat»;           !Los datos deben estar en un solo archivo concatenado
NGROUPS=2;               !Especificar el número de grupos
CASES=500;               !Especificar el tamaño de la muestral del grupo 1
VARIABLES=8;
MATRIX=RAW;
METHODS=ML;
/EQUATIONS
F1= *F10+D1;             !especificaciones de la ecuaciones del modelo de 2° orden
F2= *F10+D2;
F3= *F10+D3;
F4= *F11+D4;
F5= *F11+D5;
F6= *F11+D6;
F7= *F12+D7;
F8= *F12+D8;
F9= *F12+D9;             !A continuación se especifican de la ecuaciones del
                           modelo de 1° jorden que corresponden a la matriz A.
V1=+.8889F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;
V2=-.1111F1+.8889F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;
V3=-.1111F1-.1111F2+.8889F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;
V4=-.1111F1-.1111F2-.1111F3+.8889F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;
V5=-.1111F1-.1111F2-.1111F3-.1111F4+.8889F5-.1111F6-.1111F7-.1111F8-.1111F9;
V6=-.1111F1-.1111F2-.1111F3-.1111F4-.1111F5+.8889F6-.1111F7-.1111F8-.1111F9;
V7=-.1111F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6+.8889F7-.1111F8-.1111F9;
V8=-.1111F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7+.8889F8-.1111F9;
/VARIANCES
F10 TO F12=1.00;         !Las varianza de los factores es fijada para la
identificación
D1 TO D9=*;              !Las varianza de los errores son parámetros libres a
estimar
/COVARIANCE
F10,F11=*;               !La covarianza de los factores de 2° son parámetros libres
a
F10,F12=*;               ¡estimar
F11,F12=*;
/PRINT
PARAMETER=YES;
/OUTPUT
PARAMETERS; SS;          !Solicitar la solución estandarizada

STANDARD ERRORS;         !y los errores de estimación
/TECHNICAL
ITR=250;
/TITLE                    ¡Especificaciones para el siguiente grupo

```

Sintaxis EQS para análisis de invarianza con datos ipsatizados entre dos grupo utilizando el método DE, 3 factores, 3 variables por factor- grupo 2;

/SPECIFICATIONS

!Se suprimen las especificaciones generales y sólo se mantienen las específicas del grupo 2
CASES=300; !Especificar el tamaño de la muestral del grupo 2

VARIABLES=8;

/EQUATIONS

F1= *F10+D1;

F2= *F10+D2;

F3= *F10+D3;

F4= *F11+D4;

F5= *F11+D5;

F6= *F11+D6;

F7= *F12+D7;

F8= *F12+D8;

F9= *F12+D9;

V1= +.8889F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;

V2= -.1111F1+.8889F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;

V3= -.1111F1-.1111F2+.8889F3-.1111F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;

V4= -.1111F1-.1111F2-.1111F3+.8889F4-.1111F5-.1111F6-.1111F7-.1111F8-.1111F9;

V5= -.1111F1-.1111F2-.1111F3-.1111F4+.8889F5-.1111F6-.1111F7-.1111F8-.1111F9;

V6= -.1111F1-.1111F2-.1111F3-.1111F4-.1111F5+.8889F6-.1111F7-.1111F8-.1111F9;

V7= -.1111F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6+.8889F7-.1111F8-.1111F9;

V8= -.1111F1-.1111F2-.1111F3-.1111F4-.1111F5-.1111F6-.1111F7+.8889F8-.1111F9;

/VARIANCES

F10 TO F12=1.00;

!La varianza de los factores es fijada para la identificación

D1 TO D9=*;

/COVARIANCE

F10,F11=*;

F10,F12=*;

F11,F12=*;

/CONSTRAINTS

!Finalmente se especifican las restricciones en los parámetros.

(1,F1,F10)=(2,F1,F10);
factores

!Las restricciones del nivel métrico se realizan sobre los

(1,F2,F10)=(2,F2,F10);
los

!de 2° orden del método DE, ya que estos corresponden a

(1,F3,F10)=(2,F3,F10);

!factores de 1°orden del modelo normativo.

(1,F4,F11)=(2,F4,F11);

!Suprimiendo el comando CONSTRAINTS se obtienen la

(1,F5,F11)=(2,F5,F11);

!solución del nivel configural

(1,F6,F11)=(2,F6,F11);

(1,F7,F12)=(2,F7,F12);

(1,F8,F12)=(2,F8,F12);

(1,F9,F12)=(2,F9,F12);

/END.

ANEXO 2

Sintaxis EQS para la estimación del modelo del Inventario de Estilos de Aprendizaje de KOLB (1985) para la versión *ipsatizada*.

```

/TITLE
  Análisis de invarianza factorial para el test de KOLB (versión reducida)
  por sexo. Estudio empírico.
/SPECIFICATIONS
  DATA='*.dat';
  VARIABLES=15; GROUP=2; CASES=239;
  METHOD= ML, ROBUST; MATRIX=RAW;
/EQUATIONS
  F1= *F17+D1;
  F2= *F17+D2;
  F3= *F17+D3;
  F4= *F17+D4;
  F5= *F18+D5;
  F6= *F18+D6;
  F7= *F18+D7;
  F8= *F18+D8;
  F9= *F19+D9;
  F10=*F19+D10;
  F11=*F19+D11;
  F12=*F19+D12;
  F13=*F20+D13;
  F14=*F20+D14;
  F15=*F20+D15;
  F16=*F20+D16;
  V1=+.9375F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V2=-.0625F1+.9375F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V3=-.0625F1-.0625F2+.9375F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V4=-.0625F1-.0625F2-.0625F3+.9375F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V5=-.0625F1-.0625F2-.0625F3-.0625F4+.9375F5-.0625F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V6=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5+.9375F6-.0625F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V7=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6+.9375F7-.0625F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V8=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7+.9375F8-.0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V9=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8+.9375F9
    -.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V10=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
    .0625F9
    +.9375F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V11=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
    .0625F9
    -.0625F10+.9375F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V12=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
    .0625F9
    -.0625F10-.0625F11+.9375F12-.0625F13-.0625F14-.0625F15-.0625F16;
  V13=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
    .0625F9
    -.0625F10-.0625F11-.0625F12+.9375F13-.0625F14-.0625F15-.0625F16;
  V14=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
    .0625F9
    -.0625F10-.0625F11-.0625F12-.0625F13+.9375F14-.0625F15-.0625F16;

```

```
V15=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14+.9375F15-.0625F16;
/VARIANCES
F17 TO F20= 1.00;
D1 TO D16= *;
/COVARIANCES
F17,F18=*;
F17,F19=*;
F17,F20=*;
F18,F19=*;
F18,F20=*;
F19,F20=*;
/PRINT
FIT=ALL;
TABLE=EQUATION;
/TECH
ITR=200;
/END
/TITLE
ipsatizados
/SPECIFICATIONS
CASES=308; VARIABLES=16; MATRIX=RAW;
/EQUATIONS
/EQUATIONS
F1= *F17+D1;
F2= *F17+D2;
F3= *F17+D3;
F4= *F17+D4;
F5= *F18+D5;
F6= *F18+D6;
F7= *F18+D7;
F8= *F18+D8;
F9= *F19+D9;
F10=*F19+D10;
F11=*F19+D11;
F12=*F19+D12;
F13=*F20+D13;
F14=*F20+D14;
F15=*F20+D15;
F16=*F20+D16;
V1=+.9375F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V2=-.0625F1+.9375F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V3=-.0625F1-.0625F2+.9375F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V4=-.0625F1-.0625F2-.0625F3+.9375F4-.0625F5-.0625F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V5=-.0625F1-.0625F2-.0625F3-.0625F4+.9375F5-.0625F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V6=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5+.9375F6-.0625F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V7=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6+.9375F7-.0625F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V8=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7+.9375F8-.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V9=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8+.9375F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V10=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
+.9375F10-.0625F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
```

AFC de variables ipsativas e invarianza factorial

```
V11=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10+.9375F11-.0625F12-.0625F13-.0625F14-.0625F15-.0625F16;
V12=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10-.0625F11+.9375F12-.0625F13-.0625F14-.0625F15-.0625F16;
V13=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10-.0625F11-.0625F12+.9375F13-.0625F14-.0625F15-.0625F16;
V14=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13+.9375F14-.0625F15-.0625F16;
V15=-.0625F1-.0625F2-.0625F3-.0625F4-.0625F5-.0625F6-.0625F7-.0625F8-
.0625F9
-.0625F10-.0625F11-.0625F12-.0625F13-.0625F14+.9375F15-.0625F16;
/VARIANCES
F17 TO F20= 1.00;
D1 TO D16= *;
/COVARIANCES
F17,F18=*;
F17,F19=*;
F17,F20=*;
F18,F19=*;
F18,F20=*;
F19,F20=*;
!/CONSTRAINTS
!SET=GVF;
/LMTEST
/WTEST
!/CONSTRAINTS
!(1,F1,F17)=(2,F1,F17);
!(1,F2,F17)=(2,F2,F17);
!(1,F3,F17)=(2,F3,F17);
!(1,F4,F17)=(2,F4,F17);
!(1,F5,F18)=(2,F5,F18);
!(1,F6,F18)=(2,F6,F18);
!(1,F7,F18)=(2,F7,F18);
!(1,F8,F18)=(2,F8,F18);
!(1,F9,F19)=(2,F9,F19);
!(1,F10,F19)=(2,F10,F19);
!(1,F11,F19)=(2,F11,F19);
!(1,F12,F19)=(2,F12,F19);
!(1,F13,F20)=(2,F13,F20);
!(1,F14,F20)=(2,F14,F20);
!(1,F15,F20)=(2,F15,F20);
!(1,F16,F20)=(2,F16,F20);
/END
```

ANEXO 3

Sintaxis MPLUS para la estimación del modelo TRI multidimensional de Brown y Maydeu-Olivares (2011) del Inventario de Estilos de Aprendizaje de KOLB (1985) para la versión *ipsatizada*.

```
TITLE: Modelo de respuesta forzada de 4 factores presentado e n cuatro bloques de 4 ítems por factor
DATA: FILE IS 'MCFA_sexo.dat';
VARIABLE:
! Se asume que el archive de datos contienen solos los ítems en codificación binaria. Cualquier otra
variable !deber ser especificada.
NAMES ARE
!Se especifican los 18 comparaciones binarias y la variable de agrupación
i1i2 i1i3 i1i4 i2i3 i2i4 i3i4
i5i6 i5i7 i5i8 i6i7 i6i8 i7i8
i9i10 i9i11 i9i12 i10i11 i10i12 i11i12
i13i14 i13i15 i13i16 i14i15 i14i16 i15i16 g;
GROUPING IS g (1=male 2=female);
USEVARIABLES ARE i1i2-i15i16;
CATEGORICAL ARE ALL;
ANALYSIS:
ESTIMATOR = WLSMV;
PARAMETERIZATION = theta;
MODEL:
Trait1 BY
i1i2*1 (L1)
i1i3*1 (L1)
i1i4*1 (L1)
i5i6*1 (L5)
i5i7*1 (L5)
i5i8*1 (L5)
i9i10*1 (L9)
i9i11*1 (L9)
i9i12*1 (L9)
i13i14*1 (L13)
i13i15*1 (L13)
i13i16*1 (L13);
Trait2 BY
i1i2*-1 (L2_n)
i2i3*1 (L2)
i2i4*1 (L2)
i5i6*-1 (L6_n)
i6i7*1 (L6)
i6i8*1 (L6)
i9i10*-1 (L10_n)
i10i11*1 (L10)
i10i12*1 (L10)
i13i14*-1 (L14_n)
i14i15*1 (L14)
i14i16*1 (L14);
Trait3 BY
i1i3*-1 (L3_n)
i2i3*-1 (L3_n)
```

```

i3i4*1 (L3)
i5i7*-1 (L7_n)
i6i7*-1 (L7_n)
i7i8*1 (L7)
i9i11*-1 (L11_n)
i10i11*-1 (L11_n)
i1i12*1 (L11)
i13i15*-1 (L15_n)
i14i15*-1 (L15_n)
i15i16*1 (L15);
Trait4 BY
i1i4*-1 (L4_n)
i2i4*-1 (L4_n)
i3i4*-1 (L4_n)
i5i8*-1 (L8_n)
i6i8*-1 (L8_n)
i7i8*-1 (L8_n)
i9i12*-1 (L12_n)
i10i12*-1 (L12_n)
i1i12*-1 (L12_n)
i13i16*-1 (L16_n)
i14i16*-1 (L16_n)
i15i16*-1 (L16_n);
! la varianza de los 4 factores son restrnjidos a 1
Trait1-Trait4@1;
! Las correlaciones entre los factores son parametros libres a estimar
Trait1 WITH Trait2*0 Trait3*0 Trait4*0;
Trait2 WITH Trait3*0 Trait4*0;
Trait3 WITH Trait4*0;
Especificación de las unicidades
i1i2*1 (e1e2);
i1i3*1 (e1e3);
i1i4*1 (e1e4);
i2i3*1 (e2e3);
i2i4*1 (e2e4);
i3i4*1 (e3e4);
i5i6*1 (e5e6);
i5i7*1 (e5e7);
i5i8*1 (e5e8);
i6i7*1 (e6e7);
i6i8*1 (e6e8);
i7i8*1 (e7e8);
i9i10*1 (e9e10);
i9i11*1 (e9e11);
i9i12*1 (e9e12);
i10i11*1 (e10e11);
i10i12*1 (e10e12);
i1i12*1 (e11e12);
i13i14*1 (e13e14);
i13i15*1 (e13e15);
i13i16*1 (e13e16);
i14i15*1 (e14e15);
i14i16*1 (e14e16);
i15i16*1 (e15e16);
! Las correlaciones de las unicidades son parámetros libres a estimar
ili2 WITH ili3*1 (e1);

```

ili2 WITH ili4*1 (e1);
ili2 WITH i2i3*-1 (e2_n);
ili2 WITH i2i4*-1 (e2_n);
ili3 WITH ili4*1 (e1);
ili3 WITH i2i3*1 (e3);
ili3 WITH i3i4*-1 (e3_n);
ili4 WITH i2i4*1 (e4);
ili4 WITH i3i4*1 (e4);
i2i3 WITH i2i4*1 (e2);
i2i3 WITH i3i4*-1 (e3_n);
i2i4 WITH i3i4*1 (e4);
i5i6 WITH i5i7*1 (e5);
i5i6 WITH i5i8*1 (e5);
i5i6 WITH i6i7*-1 (e6_n);
i5i6 WITH i6i8*-1 (e6_n);
i5i7 WITH i5i8*1 (e5);
i5i7 WITH i6i7*1 (e7);
i5i7 WITH i7i8*-1 (e7_n);
i5i8 WITH i6i8*1 (e8);
i5i8 WITH i7i8*1 (e8);
i6i7 WITH i6i8*1 (e6);
i6i7 WITH i7i8*-1 (e7_n);
i6i8 WITH i7i8*1 (e8);
i9i10 WITH i9i11*1 (e9);
i9i10 WITH i9i12*1 (e9);
i9i10 WITH i10i11*-1 (e10_n);
i9i10 WITH i10i12*-1 (e10_n);
i9i11 WITH i9i12*1 (e9);
i9i11 WITH i10i11*1 (e11);
i9i11 WITH i11i12*-1 (e11_n);
i9i12 WITH i10i12*1 (e12);
i9i12 WITH i11i12*1 (e12);
i10i11 WITH i10i12*1 (e10);
i10i11 WITH i11i12*-1 (e11_n);
i10i12 WITH i11i12*1 (e12);
i13i14 WITH i13i15*1 (e13);
i13i14 WITH i13i16*1 (e13);
i13i14 WITH i14i15*-1 (e14_n);
i13i14 WITH i14i16*-1 (e14_n);
i13i15 WITH i13i16*1 (e13);
i13i15 WITH i14i15*1 (e15);
i13i15 WITH i15i16*-1 (e15_n);
i13i16 WITH i14i16*1 (e16);
i13i16 WITH i15i16*1 (e16);
i14i15 WITH i14i16*1 (e14);
i14i15 WITH i15i16*-1 (e15_n);
i14i16 WITH i15i16*1 (e16);
MODEL CONSTRAINT:

!Las saturaciones factoriales relacionadas con los mismos ítems son restringidas a ser iguales en valor

¡absoluto

L2_n = -L2;

L3_n = -L3;

L6_n = -L6;

L7_n = -L7;

L10_n = -L10;

L11_n = -L11;

```
L14_n = -L14;  
L15_n = -L15;  
!L1 = L5; L5 = L9; L9 = L13;  
!L2_n = L6_n; L6_n = L10_n; L10_n = L14_n;  
!L3_n = L7_n; L7_n = L11_n; L11_n = L15_n;  
!L4_n = L8_n; L8_n = L12_n; L12_n = L16_n;  
!Las unicidades relacionadas con los mismos ítems son restringidas a ser iguales en valor absoluto  
e2_n = -e2;  
e3_n = -e3;  
e6_n = -e6;  
e7_n = -e7;  
e10_n = -e10;  
e11_n = -e11;  
e14_n = -e14;  
e15_n = -e15;  
! Las unicidades de los pares de ítems son restringidos a ser iguales a la suma de las dos unicidades  
e1e2 = e1 + e2;  
e1e3 = e1 + e3;  
e1e4 = e1 + e4;  
e2e3 = e2 + e3;  
e2e4 = e2 + e4;  
e3e4 = e3 + e4;  
e5e6 = e5 + e6;  
e5e7 = e5 + e7;  
e5e8 = e5 + e8;  
e6e7 = e6 + e7;  
e6e8 = e6 + e8;  
e7e8 = e7 + e8;  
e9e10 = e9 + e10;  
e9e11 = e9 + e11;  
e9e12 = e9 + e12;  
e10e11 = e10 + e11;  
e10e12 = e10 + e12;  
e11e12 = e11 + e12;  
e13e14 = e13 + e14;  
e13e15 = e13 + e15;  
e13e16 = e13 + e16;  
e14e15 = e14 + e15;  
e14e16 = e14 + e16;  
e15e16 = e15 + e16;  
! Fijar una unicidad por bloque para la identificación  
e1=.5;  
e5=.5;  
e9=.5;  
e13=.5;
```

ANEXO 4
Parámetros teóricos utilizados en el estudio 1.

	<i>9 ítems</i>		<i>12 ítems</i>		<i>18 ítems</i>	
	λ	ε	λ	ε	λ	ε
<i>x1</i>	1.0	1.0	1.0	1.0	1.0	1.0
<i>x2</i>	2.0	4.0	2.0	4.0	1.0	1.0
<i>x3</i>	3.0	9.0	3.0	9.0	2.0	4.0
<i>x4</i>	1.0	1.0	4.0	16.0	2.0	4.0
<i>x5</i>	2.0	4.0	1.0	1.0	3.0	9.0
<i>x6</i>	3.0	9.0	2.0	4.0	3.0	9.0
<i>x7</i>	1.0	1.0	3.0	9.0	1.0	1.0
<i>x8</i>	2.0	4.0	4.0	16.0	1.0	1.0
<i>x9</i>	3.0	9.0	1.0	1.0	2.0	4.0
<i>x10</i>			2.0	4.0	2.0	4.0
<i>x11</i>			3.0	9.0	3.0	9.0
<i>x12</i>			4.0	16.0	3.0	9.0
<i>x13</i>					1.0	1.0
<i>x14</i>					1.0	1.0
<i>x15</i>					2.0	4.0
<i>x16</i>					2.0	4.0
<i>x17</i>					3.0	9.0
<i>x18</i>					3.0	9.0

ANEXO 5

Parámetros teóricos utilizados en el estudio 2.

<i>Condición de invarianza y no invarianza homogénea</i>												
	<i>Modelo 1</i>						<i>Modelo 2</i>					
	λ	ε	λ	ε	λ	ε	λ	ε	λ	ε	λ	ε
<i>x1</i>	1.0	1.0	1.0	1.0	1.0	1.0	0.71	1.50	0.71	1.50	0.71	1.50
<i>x2</i>	2.0	4.0	2.0	4.0	1.0	1.0	1.41	6.00	1.41	6.00	0.71	1.50
<i>x3</i>	3.0	9.0	3.0	9.0	2.0	4.0	2.12	13.50	2.12	13.50	1.41	6.00
<i>x4</i>	1.0	1.0	4.0	16.0	2.0	4.0	0.71	1.50	2.83	24.00	1.41	6.00
<i>x5</i>	2.0	4.0	1.0	1.0	3.0	9.0	1.41	6.00	0.71	1.50	2.12	13.50
<i>x6</i>	3.0	9.0	2.0	4.0	3.0	9.0	2.12	13.50	1.41	6.00	2.12	13.50
<i>x7</i>	1.0	1.0	3.0	9.0	1.0	1.0	0.71	1.50	2.12	13.50	0.71	1.50
<i>x8</i>	2.0	4.0	4.0	16.0	1.0	1.0	1.41	6.00	2.83	24.00	0.71	1.50
<i>x9</i>	3.0	9.0	1.0	1.0	2.0	4.0	2.12	13.50	0.71	1.50	1.41	6.00
<i>x10</i>			2.0	4.0	2.0	4.0			1.41	6.00	1.41	6.00
<i>x11</i>			3.0	9.0	3.0	9.0			2.12	13.50	2.12	13.50
<i>x12</i>			4.0	16.0	3.0	9.0			2.83	24.00	2.12	13.50
<i>x13</i>					1.0	1.0					0.71	1.50
<i>x14</i>					1.0	1.0					0.71	1.50
<i>x15</i>					2.0	4.0					1.41	6.00
<i>x16</i>					2.0	4.0					1.41	6.00
<i>x17</i>					3.0	9.0					2.12	13.50
<i>x18</i>					3.0	9.0					2.12	13.50

<i>Condición de no invarianza mixta</i>												
	<i>Modelo 3</i>						<i>Modelo 4</i>					
	λ	ε	λ	ε	λ	ε	λ	ε	λ	ε	λ	ε
<i>x1</i>	1.00	1.00	1.00	1.00	1.00	1.00	0.71	1.50	0.71	1.50	0.71	1.50
<i>x2</i>	1.41	6.00	1.41	6.00	0.71	1.50	2.00	4.00	2.00	4.00	1.00	1.00
<i>x3</i>	3.00	9.00	3.00	9.00	2.00	4.00	2.12	13.50	2.12	13.50	1.41	6.00
<i>x4</i>	0.71	1.50	2.83	24.00	1.41	6.00	1.00	1.00	4.00	16.00	2.00	4.00
<i>x5</i>	2.00	4.00	1.00	1.00	3.00	9.00	1.41	6.00	0.71	1.50	2.12	13.50
<i>x6</i>	2.12	13.50	1.41	6.00	2.12	13.50	3.00	9.00	2.00	4.00	3.00	9.00
<i>x7</i>	1.00	1.00	3.00	9.00	1.00	1.00	0.71	1.50	2.12	13.50	0.71	1.50
<i>x8</i>	1.41	6.00	2.83	24.00	0.71	1.50	2.00	4.00	4.00	16.00	1.00	1.00
<i>x9</i>	3.00	9.00	1.00	1.00	2.00	4.00	2.12	13.50	0.71	1.50	1.41	6.00
<i>x10</i>			2.00	4.00	1.41	6.00			2.00	4.00	2.00	4.00
<i>x11</i>			2.12	13.50	3.00	9.00			2.12	13.50	2.12	13.50
<i>x12</i>			4.00	16.00	2.12	13.50			4.00	16.00	3.00	9.00
<i>x13</i>					1.00	1.00					0.71	1.50
<i>x14</i>					0.71	1.50					1.00	1.00
<i>x15</i>					2.00	4.00					1.41	6.00
<i>x16</i>					1.41	6.00					2.00	4.00
<i>x17</i>					3.00	9.00					2.12	13.50
<i>x18</i>					2.12	13.50					3.00	9.00

Nota. En la condición de invarianza el grupo 1 y el grupo 2 han sido simulados con los valores del modelo 1; para la condición de no invarianza homogénea, el modelo 1 y 2 corresponde al grupo 1 y 2, respectivamente; finalmente para la condición de no invarianza mixta los modelos 3 y 4 corresponden a los grupos 1 y 2 respectivamente

ANEXO 6

Instrumento utilizado en el estudio empírico. Inventario de Estilos de Aprendizaje de Kolb (1985). Versión reducida.

Estimado/a participante:

El presente cuestionario es parte de un estudio que tiene como objetivo conocer la forma como los estudiantes aprenden. Ya que es una encuesta de opinión **no existen respuestas correctas o erróneas**, por lo que te pedimos que leas atentamente cada pregunta y seas **lo más sincera/o posible**. Es una encuesta anónima debido a que no nos interesa realizar análisis individuales; los análisis serán realizados de forma grupal y sólo tendrán propósitos investigativos. Te agradecemos desde ya tu colaboración, la cual será de gran ayuda para nuestra investigación.

Sexo <input type="checkbox"/> Masculino <input type="checkbox"/> Femenino	Carrera: _____
Edad: <input style="width: 50px;" type="text"/>	

A continuación te presentamos una serie de enunciados cada uno seguido de cuatro alternativas de respuesta. Te pedimos que leas atentamente cada enunciado y completes la frase ordenando las alternativas de respuestas desde la que **más te representa (4) a la que menos te representa (1)**.

La fruta que más me gusta es...			
<input style="width: 20px;" type="text" value="1"/> Manzana	<input style="width: 20px;" type="text" value="3"/> Pera	<input style="width: 20px;" type="text" value="2"/> Plátano	<input style="width: 20px;" type="text" value="4"/> Naranja

En el ejemplo: la fruta que **MÁS TE GUSTA** [4] **ES LA NARANJA**, luego la siguiente qué te gusta es la [3] PERA, luego el [2] PLÁTANO, y la que **MENOS TE GUSTA** [1] la MANZANA.

A continuación debes responder a la pregunta: **¿CÓMO APRENDO MEJOR?:**

1. Aprendo mejor cuando... <div style="display: flex; justify-content: space-between; margin-top: 10px;"> <div><input type="checkbox"/> confío en mi intuición y sentimientos</div> <div><input type="checkbox"/> escucho y observo cuidadosamente</div> <div><input type="checkbox"/> confío en el pensamiento lógico</div> <div><input type="checkbox"/> trabajo duro para lograr hacer las cosas</div> </div>
2. Yo aprendo... <div style="display: flex; justify-content: space-between; margin-top: 10px;"> <div><input type="checkbox"/> sintiendo</div> <div><input type="checkbox"/> Observando</div> <div><input type="checkbox"/> pensando</div> <div><input type="checkbox"/> haciendo</div> </div>
3. Cuando estoy aprendiendo... <div style="display: flex; justify-content: space-between; margin-top: 10px;"> <div><input type="checkbox"/> soy una persona intuitiva</div> <div><input type="checkbox"/> soy una observadora</div> <div><input type="checkbox"/> soy una persona lógica</div> <div><input type="checkbox"/> soy una persona activa</div> </div>

4. Yo aprendo mejor...

☐ de las relaciones con las personas ☐ de la observación ☐ de las teorías racionales ☐ de la oportunidad de probar y practicar

A continuación te presentamos una serie de enunciados relacionados con tu forma de aprender. Te pedimos que leas atentamente cada enunciado e indiques el grado de acuerdo con cada uno de ellos. Para ello dispones de una escala de 1 a 5 puntos donde, como está indicado en el encabezamiento de la tabla, 1 significa “totalmente en desacuerdo” y 5 “totalmente de acuerdo”. Recuerda que no hay respuestas buenas o malas y lo que nos interesa es sólo tu opinión sincera. **SÓLO DEBES MARCA UNA ALTERNATIVA.**

	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
1. Aprendo mejor cuando confío en mi intuición	1	2	3	4	5
2. Yo aprendo observando	1	2	3	4	5
3. Cuando estoy aprendiendo soy una persona lógica	1	2	3	4	5
4. Yo aprendo mejor de las oportunidades de probar y practicar	1	2	3	4	5
5. Aprendo mejor cuando escucho y observo cuidadosamente	1	2	3	4	5
6. Yo aprendo pensando	1	2	3	4	5
7. Cuando estoy aprendiendo soy una persona activa	1	2	3	4	5
8. Yo aprendo mejor de las relaciones de las personas	1	2	3	4	5
9. Aprendo mejor cuando confío en el pensamiento lógico	1	2	3	4	5
10. Yo aprendo haciendo	1	2	3	4	5
11. Cuando estoy aprendiendo soy una persona intuitiva	1	2	3	4	5
12. Yo aprendo mejor de la observación	1	2	3	4	5
13. Aprendo mejor cuando trabajo duro para lograr hacer las cosas	1	2	3	4	5
14. Yo aprendo sintiendo	1	2	3	4	5
15. Cuando estoy aprendiendo soy una persona observadora	1	2	3	4	5
16. Yo aprendo mejor de las teorías racionales	1	2	3	4	5

Muchas gracias por tu colaboración.